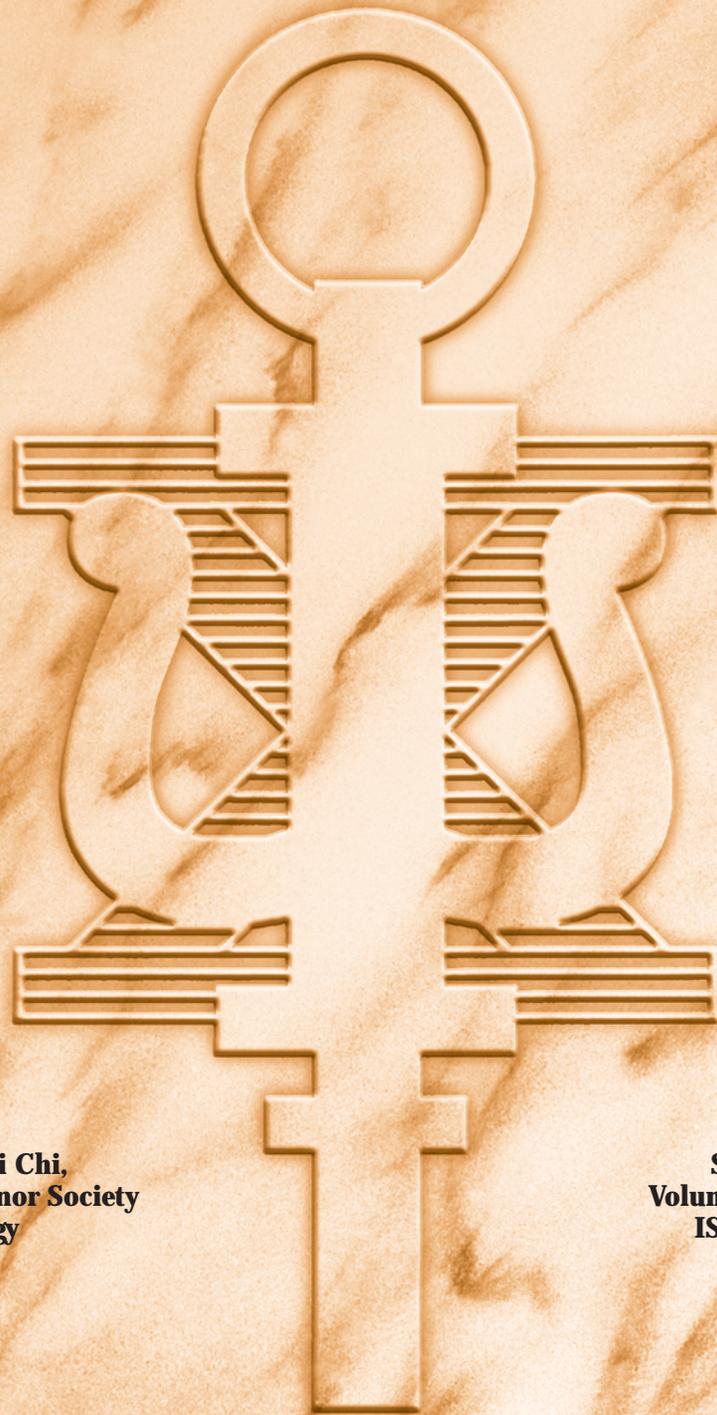


PSI CHI

Journal of Undergraduate Research



**Published by Psi Chi,
The International Honor Society
in Psychology**

**Summer 2010
Volume 15 Number 2
ISSN 1089-4136**

A large, light gray watermark of the Psi Chi logo is centered on the page. The logo features a central shield with a scale of justice, flanked by two figures holding a banner. Above the shield is a circular emblem with a caduceus-like symbol. The text 'PSI CHI' is prominently displayed at the top of the logo.

PSI CHI

Journal of Undergraduate Research

Interim Editor

Martha S. Zlokovich, PhD
Psi Chi Central Office
P.O. Box 709
Chattanooga, TN 37401-0709
Telephone: +1-573-651-9139
Fax: +1-423-265-1529
E-mail: martha.zlokovich@psichi.org

Editor

Randolph A. Smith, PhD
Department of Psychology
Lamar University
P.O. Box 10036
Beaumont, TX 77710
Telephone: +1-409-880-8285
Fax: +1-409-880-1779
E-mail: randolph.smith@psichi.org

Managing Editor

Susan Iles
Publication and Subscription Office
Psi Chi Central Office
P.O. Box 709
Chattanooga, TN 37401-0709
Telephone: +1-423-756-2044
Fax: +1-423-265-1529
E-mail: susan.iles@psichi.org



PSI CHI

Journal of Undergraduate Research

Published quarterly by Psi Chi, The International Honor Society in Psychology

SUMMER 2010 VOLUME 15, NUMBER 2

- 43 **Are You Aware? Context Effects on Emotional Ratings***
Jessica Rae Cabral and Connie Shears[†], Chapman University
- 50 **Exploring Learning Measures During Training on a Truck-Dispatcher Task***
Dorothy Carter, Debra Steele-Johnson[†], and Dragana Claffin, Wright State University
- 60 **College Students' Perceptions of Alcohol Use in Society and in Entertainment Television***
Lindsay G. Cook and Richard Jackson Harris[†], Kansas State University
- 72 **Sex Differences in 2D:4D Ratios, Aggression, and Spatial/Verbal Abilities***
Patrick Cooper and Robert Matchock[†], The Pennsylvania State University-Altoona Campus
- 76 **Influence of Name Cues to Ethnicity on Judgments of Guilt and Sentencing***
Jessica L. Davison, Danielle I. Schneider, and Sheila Brownlow[†], Catawba College
- 84 **Could Attempts to Reduce the Effects of Inadmissible Evidence Go Too Far?***
Justin Joseph, John Carroll University
- 91 **Strong Body, Strong Mind: Self-Efficacy as a Function of Exercise***
Mary Beth Miller, Westminster College
- 97 **An Eye for an Eye: Examining Personality-Based Predictors of Trait Vengefulness**
Sean E. Wisnieski, Shepherd University
- 106 **A Preliminary Examination of the Effects of Behavioral Descriptions on On-Task Behavior**
Kristin B. Lemaster, Stephanie M. Wagner, Ashley B. Tempel, and Cheryl B. McNeil[†], West Virginia University
- 114 **Self-Silencing Attitudes and Behaviors in Adolescent Couples: Links With Relationship Satisfaction**
Tiffany Michaelson and Renee V. Galliher[†], Utah State University

[†] Faculty mentors

* These articles were accepted during the editorial term of Martha S. Zlokovich.



PSI CHI

Summer 2010 | Volume 15, Number 2

Journal of Undergraduate Research

Published quarterly by Psi Chi, The International Honor Society in Psychology

For more information about Psi Chi, contact:
Psi Chi Central Office
P.O. Box 709
Chattanooga, TN 37401-0709
(423) 756-2044
www.psichi.org

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 men and women who are making the study of psychology one of their major interests and who meet the minimum qualifications. 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 in the U.S., Canada, and Ireland. The Psi Chi Central Office is located in Chattanooga, Tennessee. A Board of Directors, composed of psychologists 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—one immediate and visibly rewarding to the individual member, the other slower and more difficult to accomplish, but offering greater rewards in the long run. The first of these is the Society's obligation to provide academic recognition to its inductees by the mere fact of membership. The second goal is the obligation of each of the Society's local chapters to nurture the spark of that accomplishment by offering a climate congenial to its creative development. For example, the chapters make active attempts to nourish and stimulate professional growth through programs designed to augment and enhance the regular curriculum and to provide practical experience and fellowship through affiliation with the chapter. In addition, the organization provides programs to help achieve these goals including regional and Society conventions, research award and grant competitions, certificate recognition programs, chapter awards, and Society service projects.

Journal Purpose Statement

The twofold purpose of the *Psi Chi Journal of Undergraduate Research* is to foster and reward the scholarly efforts of undergraduate psychology students as well as to provide them with a valuable learning experience. The articles published in this journal represent primarily the work of the undergraduate student(s). Faculty mentors, who deserve recognition, are identified by an asterisk next to their name or on a separate byline.

Because the articles in this journal are primarily the work of undergraduate students, the reader should bear in mind that: (1) the studies are possibly less complex in design, scope, or sampling than professional publications and (2) the studies are not limited to significant findings. The basis for accepting papers for publication is the agreement among three professional reviewers that the project, hypothesis, and design are well researched and conceived for someone with an undergraduate level of competence and experience.

Instructions for Contributors

The *Psi Chi Journal of Undergraduate Research* encourages undergraduate students to submit manuscripts for consideration. Submissions are accepted for review on an ongoing basis. Although manuscripts are limited to empirical research, they may cover any topical area in the psychological sciences.

1. The primary author of a submitted manuscript must be an undergraduate student who is a member of Psi Chi. Manuscripts from graduate students will be accepted only if the work was completed as an undergraduate student and not more than 6 months has passed since graduation. Additional authors other than the primary author may include non-Psi Chi students as well as the faculty mentor or supervisor. Membership verification information (member ID number) for the primary author must be included.
2. Only original manuscripts (not published or accepted for publication elsewhere) will be accepted.
3. All manuscripts must be prepared according to the *Publication Manual of the American Psychological Association* (6th ed.).
4. What to submit:
 - a. A Microsoft® Word electronic copy of the complete manuscript with figures, tables, and charts generated in either Word or Excel. Any scanned images or illustrations must be at least 600 dpi resolution. Make sure that identifying names, affiliations, etc. appear only on the title page and nowhere else on the manuscript (i.e., manuscripts should be reasonably free of clues to the identity of the authors). Footnotes that identify the author(s) should appear on a separate page.
 - b. An email address so that receipt of your manuscript can be acknowledged.
 - c. A sponsoring statement from the faculty mentor who attests:
 - (1) that the research adhered to APA ethical standards;
 - (2) that the mentor has read and critiqued the manuscript on content, method, APA style, grammar, and overall presentation; and
 - (3) that the planning, execution, and writing of the manuscript represent primarily the work of the undergraduate student.

Submit all electronic files to: psichijournal@psichi.org

The *Psi Chi Journal of Undergraduate Research* (ISSN 1089-4136) 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, PO Box 709, Chattanooga, TN 37401-0709; telephone +1-423-756-2044; fax +1-423-265-1529; e-mail journal@psichi.org. Printed in the USA. 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 its publications *Eye on Psi Chi* or *Psi Chi Journal of Undergraduate Research*. The society does not wish to appear to endorse any particular products or services. Copyright 2010 by Psi Chi, The International Honor Society in Psychology. Postmaster: Send address changes to Psi Chi Journal of Undergraduate Research, PO Box 709, Chattanooga, TN 37401-0709.

Are You Aware? Context Effects on Emotional Ratings

JESSICA RAE CABRAL

CONNIE SHEARS*

Chapman University

This study examined the effects of context on emotional ratings using standardized ratings to form 3 levels of images. Participants, shown only 1 level of context, rated each image using a Likert Scale. We proposed that if context effects influence emotional ratings, then ratings would differ by weighted stimulus contexts. Results indicated that context effects altered how participants rated images. However, when participants were informed to be aware of context effects, ratings were closer to standardized ratings. These results show that emotional responses are influenced by context and further, that awareness of context can overcome this influence.

Context, defined as the environment surrounding a stimulus event, is a phenomenon present across all aspects of life (VandenBos, 2006). Context has a direct impact on how we function on a daily basis because memory, learning, judgment, and other cognitive processes are all subject to the effects of context through verbal, written, and visual information (VandenBos). Therefore, the possibility that context effects impact people's perception of their external environment deserves critical attention. Individuals cannot accurately assess incoming information unless they are aware that context influences their perception and responses. The current study investigates the potential influence of context effects on emotional judgments when people are unaware of contexts, compared to when people are aware of contexts, using the International Affective Picture System database (IAPS; Bradley, Greenwald, Petry, & Lang, 1992). We propose that if context effects influence emotional ratings, then participants' ratings will differ from the IAPS standardized ratings.

As long ago as 1962, Schachter and Singer hypothesized that emotions are a result of cognitive and physiological states. These investigators manipulated physiological state by injecting participants with either epinephrine or a placebo. Some participants were informed of side effects, other participants were not informed, and others were misinformed. All participants were then placed in a waiting room with a confederate who behaved either in an angry or euphoric manner, in order to create the "context" for the participants' responses. Results indicated that having information about potential side effects moderated participants'

emotional responses, suggesting that an informed person will be able to make more appropriate responses or decisions than an uninformed person. While these findings underscore the importance of context effects, the methodology has been widely questioned. Manstead (1979) attempted to replicate the Schachter and Singer (1962) study and found conflicting results. Nevertheless, these studies provide support for the importance of context and resulting emotions.

More recently, Dahlén (2005) examined the influence of context effects presented through the media and the resulting influences on consumer choices. Specifically, Dahlén investigated the level of creativity used to present an implicit message. Creative advertisements were defined as having new tactics and presentation styles. Traditional advertisements were defined as using tactics common in advertising. Dahlén hypothesized that as the creativity of an advertisement presentation increases, the more the message will be remembered. In addition, Dahlén also expected that there would be an increase in perceived credibility and in positive responses. To test the hypotheses, Dahlén presented media advertisements both creatively and traditionally. Dahlén also expected that there would be an increase in perceived credibility and in positive responses. Advertisements in both settings were then compared for their effectiveness and influence based on participants' reactions. The results of this study confirmed that creative ads led the consumers to make positive associations with the target brand, and they were more likely to select that brand. Traditional ads were less likely to lead con-

* Faculty mentor

sumers to make positive associations with or select those brands. These results suggest that the context (creative or traditional) of advertisements may influence the affective state of an individual, raising the question of whether context effects can alter an individual's affective response to emotional images.

Correspondingly, the contexts in which options are presented appear to influence people's choices regarding those options. That is, variables such as presenting people with options, the number of options given, the order in which the options are presented, and the context in which the options are given, can be quite influential. A person's decision can be manipulated by how information is compared and contrasted, such as the price and quality of products. Hamilton (2003) examined the influence of context effects on interactive decision making. An example would be comparing products with trade-offs between the price and level of quality. Hamilton claimed a person's choices can be altered by providing alternative choices, regardless of whether or not a person actually chooses other options. Participants were presented with two scenarios: one involved making a choice from a restaurant menu and the other involved making a choice between different appliances for a new apartment. Hamilton found that people who were informed that the intention of the study was to persuade them with the use of context effects were more involved and articulate in their decision-making than people who were not informed of this intention. These findings suggest that if participants are informed of potential context effects, they will respond differently than if they are not informed.

Additionally, survey methodologists have noted that the structure and context of questions presented first on a survey will impact people's responses to the questions that follow (Hair, 2005). Hair identified the involvement of context effects in the production and implementation of survey questions. Statements were presented in two versions in an attempt to provide context for the participants' responses. One version favored animal rights (pro-version), the other was against animal rights (anti-version). The target question asked participants to indicate the degree to which they agreed or disagreed with the statement: *Animals should have the same rights as human beings*. The results indicated that participants responded in accordance to the context they received, indicating that context influences an individual's response to a question. If the participants received the context that favored animal rights (pro-version), then they agreed more with the target statement: *Animals should have the same rights as human beings*. If the participants received the context that was against animal rights (anti-version), then they disagreed more with the target statement. The results of

this experiment raised the question of whether image context will influence participants' ratings to the images of the whole set.

Celuch, Slama, and Schaffenacker (1997) investigated whether some people are more sensitive than others to situational social cues in advertisements. A situational social cue is a stimulus presented in a social setting that serves to guide behavior. The social cues were deemed either appropriate (consistent with social norms) or inappropriate (inconsistent with social norms). Socially appropriate cues were neutral, such as ads for shampoo or cereals, socially inappropriate cues were for cigarettes or contraceptives. The authors expected that scores based on the Concern for Social Appropriateness Scale (CFA) would measure people's sensitivity to social context cues and their likelihood of acting on those cues (Celuch et al.). The results indicated that people who scored higher on the CFA were more sensitive to inappropriate social context cues than people who scored lower on the CFA (Celuch et al.). Thus, in response to socially inappropriate cues people who are aware of social context cues respond differently than people who are less aware of context cues. Our intention here is to investigate whether awareness of context allows decision makers to override the influence of context.

Prior to being influenced by context, people process incoming social cues. Valdesolo and DeSteno (2006) noted that the cognitive processes individuals use to process incoming social cues may be influenced by various factors, such as moral judgments, emotions, logic, and the current social stimuli (Valdesolo and DeSteno, 2006). These possible influences are typically in effect simultaneously, agree with one another, and lead the individual to come to one conclusion. However, it is also possible that the influential factors present during cognitive processing disagree with each other, and will lead an individual to different conclusions. When this happens, it can be unclear which conclusion is correct. Valdesolo and DeSteno (2006) hypothesized that positive feelings created by the context of information, at the time a judgment is made, might reduce the perceived negativity (aversion signal) of any possible moral conflict, therefore, increasing utilitarian responses. A utilitarian response is one that is practical and benefits the greatest number of people.

To study this idea, Valdesolo and DeSteno (2006) presented 79 participants scenarios using video clips. One of the clips depicted a positive affect, the other presented a neutral affect. Participants were then asked to report their affective state in response to the presented situation. They found that participants who were presented a positive clip reported a more positive affective state and had a greater chance of selecting the

utilitarian response than those presented with a neutral clip. These findings demonstrate that judgments are not only based on the information being processed, but also on the affective characteristics of the context of that information (Valdesolo and DeSteno, 2006). This suggests that context affects the cognitive processing of emotions and we suggest the processing of emotions can be manipulated by controlling context. Experiment 1 tested the influence context effects have on participants' responses to emotional images. We proposed that if context effects influence emotional ratings, then ratings will differ by weighted stimulus context. Furthermore, if presented with more negative images, then ratings of positive and neutral images will be more positively rated than the standard rating (the ratings of the images were standardized for valence). If presented with more positive images, then ratings of negative and neutral images will be more negatively rated than the standard rating. Lastly, if presented with an equal number of positive, negative, and neutral images, then the ratings of all of the images will be similar to the standardized ratings.

Experiment 1

Method

Participants. Ninety-six students (76 women and 20 men) from a small western university volunteered to participate. All participants were at least 18 years of age and signed informed consent forms, which informed them of their right to withdraw at anytime. All participants either received course credit or were paid \$5 each.

Materials. A total of 150 standardized images acquired from the International Affective Picture System (IAPS) database were used in the experiment (Bradley et al., 1992). IAPS is an extensive set of images that have standardized ratings for valence. The ratings ranged from 1 (*most negative*) to 9 (*most positive*). For example, a basket of puppies has an average rating of 8, a coffee cup on a table has an average rating of 5, and a shark baring its teeth has an average rating of 2.

In this experiment, context effects were created by the composition of three levels of affective images: positive, negative, and neutral. The images were organized according to their standardized ratings. Based on standardized ratings, images with ratings of 1 to 3 were included in our negative context, images with ratings of 4 to 6 were included in our neutral context, and images with ratings of 7 to 9 were included in our positive context.

There were 90 images per level. Context was created by manipulating the number of positive, negative, and neutral images. The positive context group had 60 positive images, 15 negative images, and 15 neutral

images. The negative context group had 60 negative images, 15 positive images, and 15 neutral images. The neutral context group had 30 positive images, 30 negative images, and 30 neutral images. The images were grouped this way because we were interested in the effect the images in the majority had on the images in the minority, and on the ratings of the entire set of images as a whole. For example, for the level of images with 60 positive images, 15 neutral images, and 15 negative images, we were interested in the effect the 60 positive images had on the ratings of both the negative and neutral images, as well as their influence on the average rating of the entire set. We showed each context to 32 participants, totaling 96 participants. We employed varying contexts to provoke positive, neutral, or negative ratings of image content.

Apparatus. A 17-inch Dell PC computer presented the stimuli. Participants used the numeric keypad to rate the images. The other keys were inaccessible and inactive during the experiment. DirectRT was the program used to format and present the stimulus (Jarvis, 2006).

Design. We examined single independent variable (context group) with three levels by measuring the participants' ratings of the presented images and their reaction times to make those ratings.

Procedure. Participants volunteered for one session. The length of each session was approximately 20 min. Before the experiment began, the volunteer read and signed an informed consent. After the necessary forms were completed, the experimenter instructed the volunteer to sit centered in front of a computer. Instructions were given verbally and also presented on the computer screen. Each participant viewed only one level of context via a computer slideshow. The participant rated each image as positive, neutral, or negative, using the 0–9 Likert scale (0 = *negative*; 9 = *positive*). When instructions were complete, the experimenter answered any questions, and then prompted the volunteer to begin a practice session. The experimenter was present during the practice session to ensure the instructions regarding how to rate the images were understood. During the practice session, the volunteer viewed and rated one positive image, one negative image, and one neutral image. At that point, the experimenter answered any further questions and then exited the room. During both the practice session and the primary experiment, the images appeared for 1,000 ms. The Likert scale appeared between the presentations of each image to reiterate the rating scale. The scale remained on the screen until the participant selected a number key to rate the image previously shown. This process of viewing and rating images continued until the participant

had rated all 90 images. At the end of the slideshow, the participant was prompted that she or he had completed the experiment.

After the participant had rated all images, the computer displayed a message that the experiment had been completed. At this time, the participants left the testing room and met with the experimenter for a debriefing and to collect the necessary paperwork, including their copy of the Informed Consent and the campus Bill of Rights (which describes the rights of study participants and is required by our Institutional Review Board).

Results and Discussion

The data was analyzed by MANOVA using the Statistical Software Package for the Social Sciences (SPSS) 15.0 version (Green and Salkind, 2000). This analysis examined participants' ratings by context group. The results supported all hypotheses. Means and standard errors are presented in Table 1.

Ratings. The main effect of context groups was statistically significant, $F(2, 95) = 46.13, p < .001$. The average rating of all 90 images within the negative-context group was 4.79 ($SE = .037$), meaning the images were evaluated more positively than the standardized ratings when embedded within the greater number of negative images. The average rating of all 90 images within the positive-context group was 4.29 ($SE = .037$), meaning the images were evaluated more negatively than the standardized ratings when embedded within the greater number of positive images. The average rating of all 90 images within the neutral-context group was 4.56 ($SE = .30$), meaning the images were evaluated similarly to the standardized ratings when embedded within the equal number of positive, negative, and neutral images (See Table 1). Thus, the data presented here are the average ratings of each level of images (positive, negative, and

neutral) as a whole. The data that follows presents the influence that the images in majority had on the images in minority within each context group.

The interaction between the context group and the standardized ratings was also significant, $F(2, 95) = 8.57, p < .001$. The data reveal that the ratings given by the participants depended on which context they received (see Figure 1). Participants in the negative context group rated positive images (8.89) higher (more positively) than standardized ratings of the same images (7.38). Participants in the positive context group rated negative images (1.1) more negatively than standardized ratings of the same images (1.7) and there was no difference between the neutral context group ratings and the standardized ratings.

FIGURE 1

Average ratings by context groups compared to standardized ratings.

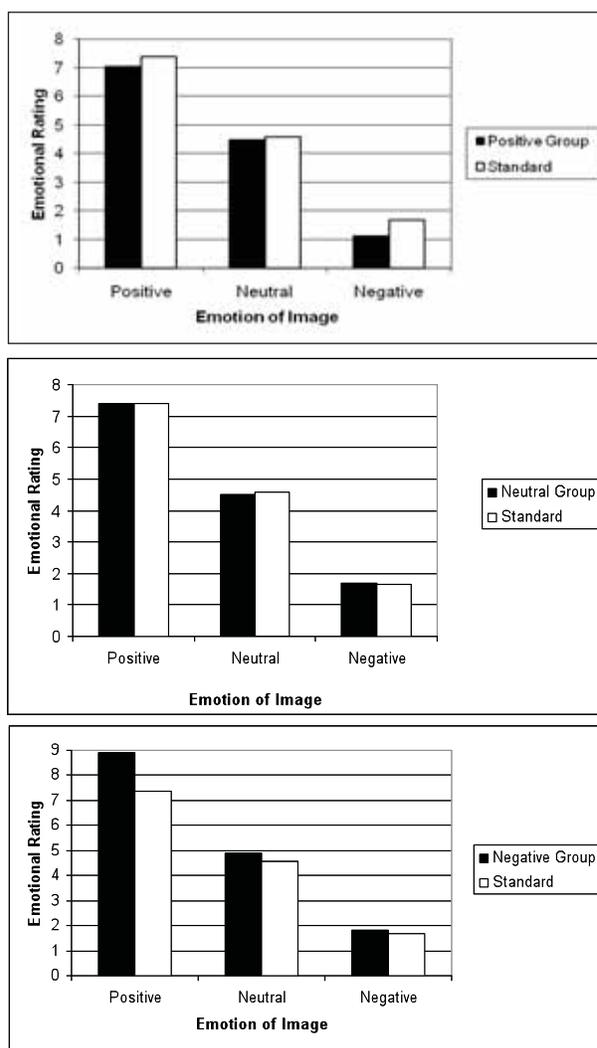


TABLE 1.

Mean Ratings and Standard Errors of Standardized Images Within Context Groups

| Type of context | Context Groups | | |
|---------------------|----------------|-------------|-------------|
| | Negative | Positive | Neutral |
| Standardized Rating | | | |
| Negative | 1.82 (1.79) | 1.38 (1.6) | 1.78 (1.49) |
| Positive | 7.63 (1.60) | 7.03 (1.46) | 7.49 (1.80) |
| Neutral | 4.90 (0.70) | 4.47 (1.41) | 4.40 (1.24) |

n = 32 in each context group.

We also examined our group ratings relative to the standardized ratings of images (see Figure 1). For standardized negative images, ratings averaged between 1 and 3, and our negative context group rated negative images more positively (1.82) relative to the positive context group (1.38), and the neutral context group (1.78). Within the positive context, the negative images were rated more negatively, relative to the negative and neutral context (1.38).

For standardized positive images, ratings averaged between 6 and 8 and our positive context group rated positive images more negatively (7.03), relative to the negative context group (7.63) and the neutral context group (7.49). The positive images within the positive context were rated more negatively, relative to the ratings of positive images within the negative and neutral contexts (7.03).

For standardized neutral images, ratings averaged between 4 and 5, and our neutral context groups rated neutral images more negatively (4.0), relative to the positive context group (4.47) and the negative context group (4.90).

Reaction time. Participants responded fastest if they were in the positive context group (1,267.75 ms). Participants in the positive and negative groups responded fastest than the neutral group perhaps because of physical arousal. Within the neutral context group, participants had the slowest reaction time average (1,698.37 ms). Such findings could have been due to the greater number of unbiased images, causing tedium. Within the negative context group, participants had a reaction time average in the middle of the positive and neutral reaction time average times (1,492.05 ms). This could have been because the image content was more engaging, thus altering their reaction time.

Experiment 1 supported our hypothesis that when participants were unaware of our manipulation of contexts, their ratings would be significantly different from standardized ratings. However, to effectively demonstrate that this finding resulted from unawareness of the context, we conducted Experiment 2 to directly test if awareness of context would move participant ratings closer to standardized ratings. We proposed that if participants are aware of context manipulations, then ratings from different contexts would not differ from standardized ratings.

Experiment 2

Method

Participants. One-hundred and twenty students (91 women and 29 men) from a small western university volunteered to participate. All participants were at least 18 years of age and signed informed consent forms, which informed them of their right to withdraw at

anytime. All participants either received course credit or were paid \$5 each.

Materials. All aspects of Experiment 2 remained exactly the same as Experiment 1 with the exception that there were two levels of awareness (unaware and aware). The unaware level provided verbal instructions identical to the verbal instructions used in Experiment 1 and included 60 participants (20 per context group). The aware level verbal instructions informed the other 60 participants (20 per context group) about possible context effects. More specifically, the participants were informed of the context level they were going to view, that the images were grouped intending to change their perception, to be aware of the effects of context, and to not let previous images influence the ratings of future images.

Design. We analyzed the hypotheses using a 3 (positive, negative, and neutral context) \times 2 (unaware and aware) between-subjects design.

Results and Discussion

Using MANOVA on the Statistical Software Package for the Social Sciences (SPSS) 15.0, the data was analyzed. We examined participants' ratings by context level and awareness level. Means and standard errors are presented in Table 2.

Ratings. The main effect of context group was statistically significant, $F(2, 119) = 34.89, p < .001$. The average rating of images within the negative context was 4.81 ($SE = .033$). The average rating of images within the positive context was 4.46 ($SE = .033$). The average rating of images within the neutral context was 4.78 ($SE = .027$). These data show that the images as a whole were evaluated differently depending on the context in which the participants viewed the images.

The main effect of standardized ratings was statistically significant, $F(2, 119) = 8154.82, p < .001$. The average rating of negative images was 2.02 ($SE = .029$). The average rating of positive images was 7.33 ($SE = .029$). The average rating of neutral images 4.70 ($SE = .035$). This data demonstrates that viewing the images within the positive or negative context groups impacted how the participants rated the images, compared to the standardized ratings.

The interaction between context awareness, context group, and standardized ratings was statistically significant, $F(2, 119) = 14.99, p < .001$. To further examine the 3-way interaction we analyzed the data via 2-way interactions by context groups. The 2-way interaction between context awareness and the negative context group was statistically significant $F(2, 119) = 23.84, p < .001$. The interaction between context awareness and the positive context group was statistically significant, $F(2, 119) = 5.53, p < .05$. The interaction

between context awareness and the neutral context group, however, was not statistically significant, which supports our hypothesis. These data provide evidence that perceptions of images presented within a negative or positive context group are impacted by whether or not the participant is unaware or aware of context (see Table 2).

Reaction time. There was a main effect of awareness (Aware vs. Unaware) in the reaction time data, $F(2, 119) = 11.40, p < .001$. Participants in the aware group were faster to respond (1495.90 ms) than participants in the unaware group (1605.03 ms). This could be a result of being more cognitively attentive to the images because the participants were informed to be aware of the effects of context.

Concerning context groups, the reaction time was statistically significant, $F(2, 119) = 3.16, p < .05$. Participants were fastest to respond if they were in the negative context group (1,491.85 ms), suggesting that the adverse images evoked an urgent response. Participants were slowest to respond if they were in the positive context group (1,591.97 ms), signifying that the favorable

images did not induce an urgent response. The reaction time for participants in the neutral context group (1,567.58 ms) was between those of the negative and positive context groups. Such findings could indicate that due to the equal number of positive, negative, and neutral images, participants were neither engaged nor repelled by the images, causing no significant effect on their reaction time.

For the standardized image ratings, the reaction time was statistically significant, $F(2, 119) = 25.72, p < .001$. Participants were fastest to respond to neutral images (1,411.84 ms), suggesting that the neutral images held no interest (positive or negative) for the participants. Participants were slowest to respond to negative images (1,699.96 ms), suggesting that participants were engaged by the adverse images. Last, the speed at which participants rated positive images was in-between that of neutral and negative images (1,539.59 ms), implying that the pleasant images were not dull or engaging.

General Discussion

Context is used by everyone, but particularly by marketing companies to effectively sway consumers to purchase their products. We have demonstrated here that even a person's emotional judgments can be swayed by context effects, and moreover, that informing a person to be aware of the potential influence of context effects allows that person to be less influenced by the context. These data are noteworthy for consumers and marketers alike, and speak to the importance of extending our understanding of context effects.

Our data from Experiment 1, where none of the participants were informed to be aware of the potential effects of context, suggest that context altered individuals' perceptions of positive, negative, and neutral images, as hypothesized. We demonstrated that when participants were presented with a context mainly composed of negative images, those participants' ratings of positive and neutral images were more positive than the standardized ratings for those images. Furthermore, when participants were presented with a context of mostly positive images, then their ratings of negative and neutral images were more negative than the standardized ratings. And in a clear demonstration of manipulating emotional context effects, we found that when participants were presented with an equal number of positive, negative, and neutral images, their ratings of all of the images were similar to the standardized ratings.

Experiment 2 investigated the influence of being aware or unaware of context effects on participants' image ratings. Again, the results supported our hypothesis. We found that if participants were aware of context

TABLE 2

Mean Ratings and Standard Errors of Standardized Images Within Context Groups and Context Awareness Groups

| Unaware Group (<i>n</i> = 60) | | | |
|--------------------------------|----------------|------------|------------|
| Type of context | Context Groups | | |
| | Negative | Positive | Neutral |
| Standardized Rating | | | |
| Negative | 1.91 (.05) | 1.81 (.09) | 2.27 (.07) |
| Positive | 7.53 (.09) | 6.79 (.05) | 7.62 (.07) |
| Neutral | 5.15 (.09) | 4.34 (.09) | 4.60 (.07) |

n = 20 per context group

| Aware Group (<i>n</i> = 60) | | | |
|------------------------------|----------------|------------|------------|
| Type of context | Context Groups | | |
| | Negative | Positive | Neutral |
| Standardized Rating | | | |
| Negative | 2.33 (.05) | 1.80 (.09) | 1.98 (.07) |
| Positive | 7.14 (.09) | 7.28 (.05) | 7.66 (.07) |
| Neutral | 4.81 (.09) | 4.76 (.09) | 4.55 (.07) |

n = 20 per context group

effects, then their ratings were similar to the standardized ratings. Specifically, participants aware of context effects rated images within the positive and negative context groups no differently than the standardized ratings.

These results indicate that informing people to be aware of the potential of context effects significantly changes their decisions about emotional ratings. This information is important to consumers because being knowledgeable about external influences, such as context, may support informed and rational decisions. We believe that further research regarding awareness of context effects would put consumers at a great advantage. Our data suggest that those who are aware of such implications would benefit because they would be less susceptible to effects of context. Furthermore, consumers would be able to use awareness as a positive cognitive tool when facing purchasing decisions.

Lastly, that context effects and context awareness alter people's emotional ratings brings to the forefront the idea that emotions can play a significant role in consumer choices. In particular, we have shown that positive, negative, and neutral contexts significantly altered the participants' emotional responses. Cohen and Pressman (2006) researched the effects of positive affect traits on health. They said that having a positive context of thought could greatly increase both mental and physical health. Their research, along with the current study, undoubtedly demonstrates that context of thought or emotion can significantly influence decision making, including consumer choices. Noticeably, much is known regarding the impact of context on perception, judgments, and consumer choices, but very little research has examined the potential of context effects on emotions. While we know that context

effects are present in verbal, written, and visual media, we know of no previous research that considers the potential of context to sway the emotions of intended consumers. Thus, the influence of context effects on emotions, which may manipulate people's perceptions, decisions, or behaviors, should be thoroughly studied. The relevance of consumer awareness of potential context effects, particularly when emotional processing is required, is an area too long neglected.

References

- Bradley, M., Greenwald, M. K., Petry, M. C., & Lang, P. J. (1992). Remembering pictures: Pleasant and arousal in memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*, 379-390.
- Celuch, K., Slama, M., & Schaffenacker, S. (1997). Concern for appropriateness and ad context effects. *Psychology & Marketing*, *14*, 19-27.
- Cohen, S., & Pressman, S. D. (2006). Positive affect and health. *Current Directions in Psychological Science*, *15*, 122-125.
- Dahlén, M. (2005). The medium as a contextual cue. *Journal of Advertising*, *34*, 89-99.
- Hair, M. (2005). Using an individual differences perspective to investigate context effects in surveys: A test of the belief-sampling model. *International Journal of Public Opinion Research*, *17*, 213-225.
- Hamilton, R. W. (2003). Why do people suggest what they do not want? Using context effects to influence others' choices. *Journal of Consumer Research*, *29*, 492-507.
- Jarvis, B. G. (2006). DirectRT (Version 2006.0.1.22) [Computer Software]. Precision Timing Software. New York, NY: Empirisoft Corporation.
- Manstead, A. S. R. (1979). A role-playing replication of Schachter and Singer's (1962) study of the cognitive and physiological determinants of emotional state. *Motivation and Emotion*, *3*, 251-264.
- Schachter, S., & Singer, J. E. (1962). Cognitive, social and physiological determinants of emotional state. *Psychological Review*, *69*, 379-399.
- Valdesolo, P., & DeSteno, D. (2006). Manipulations of emotional context shape moral judgment. *Psychological Science*, *17*, 476-477.
- VandenBos, G. R. (Ed.). (2006). *APA dictionary of psychology*. Washington DC: American Psychological Association.

Exploring Learning Measures During Training on a Truck-Dispatcher Task

DOROTHY CARTER
DEBRA STEELE-JOHNSON*
DRAGANA CLAFLIN
Wright State University

We used an inter-disciplinary approach to assess the validity of using alternative behavioral variables including task speed, information seeking, and types of errors as operational definitions of learning rather than relying on performance scores alone. Our study was a re-evaluation of a sample from a larger data set (Palumbo, 2008). Undergraduate students (N = 145) participated in a moderately difficult truck-dispatcher task (Steele-Johnson & Perlow, 1989) in which they received, processed, and shipped orders of military supplies. We categorized participants into greatest-learning versus least-learning groups based on the amount of change in their performance scores, and then compared task speed, information seeking, and types of errors between the two groups. Results indicated that task action speed and patterns of errors could be used as learning indicators for this task. Results also provided initial evidence of the efficacy of using alternative measures other than final performance score to assess learning during training.

Organizations have highlighted the importance of and spent substantial sums on training (Cascio & Aguinis, 2005; Facticeau, Dobbins, Russell, Ladd, & Kudisch, 1995). Workplace training can help employees adapt quickly to required changes in work tasks and improve their productivity (Quiñones, 1997). Thus, the focus of researchers and practitioners alike has been on ways of improving training and training outcomes. Certainly cognitive ability plays a large part in training outcomes (e.g., Hunter & Schmidt, 1996; Schmidt & Hunter, 1998), but research has shown that there are other factors, such as training type (Baldwin & Ford, 1998; Keith & Frese, 2008) or trainee characteristics (Warr & Bunce, 1995) that also influence learning during training. Researchers have operationalized learning outcomes in workplace training generally as scores on paper-and-pencil or performance tests (Arthur, Bennett, Edens & Bell, 2003). Scores from paper-and-pencil or performance tests are important because they can be directly related to the company bottom line. However, performance scores might not tell us enough. For example, it would be useful to know *what* was learned and *when*.

There is certainly precedent in the training literature for examining measures of learning other than performance scores. For example, Kraiger, Ford and Salas (1993) called for multiple outcome measures, including skill-based, cognitive, and affective outcomes, in order to more thoroughly evaluate training effective-

ness. Also, research (Austin & Bobko, 1985; Erez, 1990) has indicated there is some value in distinguishing between performance quantity outcomes (i.e., speed) and performance quality outcomes (i.e., accuracy). Moreover, when designing computerized training systems, such as intelligent tutoring systems which adapt training based on trainee errors, researchers must have access to indicators of learning that occur well before the final performance score (Steele-Johnson & Hyde, 1997). Identifying behaviors indicative of learning and then using these behaviors as signals during training to tailor training methods to the individual could increase the effectiveness of training.

When evaluating the effectiveness of workplace training, a central question is whether the training results in learning (e.g., Kraiger & Jung, 1997). However, merely calculating performance scores, as most training evaluations do, may provide an incomplete assessment of trainee learning. More specifically, it might be useful to also identify *what* was learned, that is, the content of learning. For example, individuals who make different types of errors might obtain similar performance scores, but trainers could be more effective if they knew specifically what kinds of errors trainees made. Moreover, in developmental learning research, in addition to assessing participants' end results, such as performance scores, researchers also determine

* Faculty mentor

how or when the participant learned the information needed to achieve those results (Berninger, et al., 2006; Ivkovich, Eckerman, Krasnegor, & Stanton, 2000). Additionally, information about when errors tend to be made, for example, might be helpful in designing training. That is, individuals might tend to learn some task rules before others, or it might enhance learning to make some types of errors earlier versus later in training. Thus, trainers may benefit from identifying alternative measures or indicator variables that distinguish those participants who learn more of the training content from those who learn less during training.

The purpose of the study was to examine alternative measures of learning, addressing both the content and timing of learning (early vs. late), in order to identify factors that distinguish between those who learn more and those who learn less during training. Such measures ultimately could indicate modifications that trainers or training systems could make to enhance trainee learning. We sought to determine alternative operational definitions of learning (indicator variables) that occur during training. These indicator variables could be used to differentiate early in the training process between those who are learning the material and those who may not be learning as quickly in order to modify training parameters and improve training success. We tested the validity of using several different alternative variables as indicators of learning, including task speed, information seeking, and types of errors or mistakes.

Task Speed

Hypothesis 1: Trainees' completion of initial task actions will occur more quickly with repeated task practice. As participants gain experience and task knowledge through repeated task practice, it is assumed that they will demonstrate improved skill development as evidenced by faster performance (e.g., Kraiger et al., 1993). Therefore, one factor that might be indicative of task learning is an increase in speed of task actions. Specifically, we hypothesized that with task practice participants would be more prepared to begin the task and therefore better able to complete initial task actions quickly, resulting in increased productivity and overall performance scores.

Hypothesis 2: Individuals in the greatest-learning group will demonstrate greater improvement in speed of initial task actions, relative to those in the least-learning group. Overall we expected most participants to display some improvement in the time taken to complete initial task actions after repeated practice (Hypothesis 1); however, the difference between participants' improvement in time taken to complete initial task actions may be a distinguishing variable

indicative of task learning. Comparing the improvement in the time participants take to complete an essential task action is a measure of learning based on improvement in speed that is often used in developmental and educational research (Gettinger & White, 1979; Ivkovich et al., 2000). However, researchers have rarely used this measure as an indicator of learning in I/O research (see Singer & Gaines, 1975, for an exception). For comparison purposes we operationally defined participants whose performance scores displayed the greatest increase as the greatest-learning group and those whose performance scores displayed the least increase as the least-learning group, and we omitted those individuals whose performance scores displayed moderate increases. We not only expected to see differences in the time individual participants took to complete initial task actions, we also hypothesized that participants whose performance scores showed the most improvement would demonstrate the most increase in speed of initial task actions.

Hypothesis 3: Amount of initial information seeking will be positively related to final performance score. For decades, researchers have viewed feedback as essential for learning and performance (Ilgen, Fisher & Taylor, 1979). Ashford and Cummings (1983) proposed that individuals engage in feedback-seeking behavior in order to reduce uncertainty about what goals to pursue. Participants who initially seek out information regarding task instructions and rules might be more likely to learn a task than those who do not seek out this information. Thus, another factor that may distinguish participants who display the greatest learning might be information-seeking behavior occurring during the beginning of training. The amount of initial information seeking should be positively related to final performance scores.

Hypothesis 4: Number of initial mistakes will be positively related to final performance scores. Traditionally, researchers have associated errors made during training with punishment (Skinner, 1953) and viewed errors as detrimental to learning (Bandura, 1986). However, more recent research (e.g., Dormann & Frese, 1994) has indicated that errors may be beneficial to learning in that they actually boost exploratory behavior, which in turn leads to strategy refinement and development. Researchers have shown that training which presents errors as useful or encourages errors leads to better performance on transfer or novel tasks than training which presents errors as dysfunctional or instructs learners to avoid errors (e.g., Keith & Frese, 2008). Additionally, early mistakes during training might be reflective of a more exploratory initial learning strategy. The recognition that one has made a mistake can cause further exploratory behavior.

Based on previous research demonstrating the beneficial effects of making mistakes during training (e.g., Lorenzet, Salas & Tannenbaum, 2005), we expected that trainees who made a greater number of errors early in training would gain more experience with the task and would perform better than those who did not experience these errors early on.

Hypothesis 5: Individuals in the greatest-learning group will display different types of errors relative to those in the least-learning group. Researchers in the training and motivation literature have demonstrated repeatedly that performance quantity and performance quality are not necessarily positively related (e.g., Erez & Arad, 1986). For instance, researchers have demonstrated that there can be a speed/accuracy trade-off in which increased speed of performance does not necessarily improve accuracy of performance, and vice versa (e.g., Stanton & Julian, 2002). Therefore, it may be beneficial for researchers interested in training to assess not only trainees' improvement in task speed (Hypotheses 1 and 2) but also the types of errors trainees make. Types of errors might be another factor that distinguishes participants who have achieved the greatest learning from those who have learned the least.

Hypothesis 6: Individuals in the greatest-learning group will display different patterns of change in errors during training, compared to those in the least-learning group. Different error patterns could have different effects on performance, and thus individuals demonstrate their learning through changes in the types of errors they make with task practice. This could happen in two ways. Participants who have made certain errors early in task practice may learn from those errors, make them less frequently, and go on to display improved performance. Indeed, van der Linden, Sonnentag, Frese and van Dyck (2001) observed that initial errors can sometimes cause improvements in task strategy use. On the other hand, some errors might enhance performance. For instance, ignoring certain rules in a task in favor of other more essential ones might be a task strategy used by those who have better-learned the material. We expect those in the greatest-learning group will display different patterns of changes in errors, relative to those in the least-learning group.

Method

Participants

The current study examines part of a larger data set $N = 246$ (Palumbo, 2008). The subset of the data included in the analyses described here were from 145 undergraduate students (31 men and 114 women) from a medium-sized, midwestern university. Participants ranged in age from 18 to 53 years with a mean age of 21.72 years, $SD = 4.97$. Participants had received

extra credit points that they could apply toward their course grade.

Task Description and Procedure

The larger study obtained a data set using an adaptation (Palumbo, 2008) of a well-studied, moderately difficult, truck-dispatcher task (Steele-Johnson & Perlow, 1989). All participants received task instructions and then completed five 10-min trials of the task. The task was an MS-DOS based computer simulation of duties required of a truck-dispatcher. Participants were required to receive, process, and ship orders of military parts and supplies to three areas within an Area of Responsibility. Running task time was defined as the time spanning one work week. Thus the task time began at 9:00 a.m. on Monday, and ended at 5:00 p.m. on Friday. Each hour interval in the task was equivalent to 15 s in real time, and each day in the task was equivalent to 2 min in real time. Prior to the task, participants read task instructions including seven rules concerning truck-capacity restrictions, time schedules, and delivery area restrictions.

Prior research has demonstrated that task instructions emphasizing either performance quality or performance quantity requirements can influence the speed/accuracy trade-off that is sometimes observed in individual performance (e.g., Stanton & Julian, 2002). Task instructions in the current study did not specifically stress either performance quantity (speed of shipment) or performance quality (obeying task rules); therefore, the task instructions should not have influenced one outcome more than another. Experimenters told participants that the purpose of the task was to accept incoming orders for processing, process those orders appropriately, and dispatch trucks quickly and correctly. Experimenters told participants also that they would receive points for each unit shipped and that points would be subtracted when rules were not followed.

In the truck-dispatcher task, subjects were required to ship items displayed in the service window either onto a truck or through a pick-up window (depending on item code). There were seven rules constraining how participants were to complete task actions. For instance, participants were required to limit loads on a truck to a certain level, send a truck to no more than three out of five shipment zones, and ship orders as indicated either onto a truck or to a pick-up window. See appendix for information on specific rules.

Learner Type

We created two groups of participants: those who demonstrated the greatest learning versus the least learning. This was a multi-step process. First, we computed a performance difference score by subtracting the Trial 1 score from the Trial 5 score for each par-

participant. Then we ordered the participants from those showing the greatest increases in performance scores from Trial 1 to Trial 5 to those showing the smallest increases (and including decreases) from Trial 1 to Trial 5. Finally, we identified as the greatest-learning group the 33% of participants showing the greatest increase in performance scores, and the least-learning group as the 33% of participants showing the smallest increase (including decreases). Final performance scores varied widely in both groups. There were 49 participants in the greatest-learning group (40 women and 9 men) with a mean age of 21.70 years, $SD = 5.90$ and 48 participants in the least-learning group (36 women and 12 men) with a mean age of 20.82 years, $SD = 3.22$. We omitted the 48 participants showing moderate increases (the middle 33%).

Composite Task Performance

We operationally defined composite task performance as participants' performance scores in each trial. Participants received 5 points for each unit of office equipment correctly shipped and lost 10 points for each rule violation.

Subtask Performance

We operationally defined the subtask performance as the number of times each task action was completed during a trial. These task actions are listed below. Although not all actions resulted in points added or subtracted from the performance score, all actions required time to complete, which ultimately affected performance score.

Shipped truck. When participants dispatched one of the three trucks, and the truck was permitted to leave (did not exceed truck-capacity by over 20%), the computer software recorded the truck shipment and time associated with it. We operationally defined the subtask performance for this action as the time (seconds) taken to dispatch the first truck during a task trial.

Sent out pick-up. The computer software recorded the time at which participants placed pick-up orders in the pick-up window. We operationally defined the subtask performance for this action as the time (seconds) participants took to dispatch the first pick-up order during a trial.

Rule violations. The computer software recorded each time a participant violated one of the seven rules (see Appendix). We operationally defined the subtask performance for this action as the number of rule violations (both for individual rule violations and for total number of violations) made during a trial.

Viewing of Rules 1 through 7. At any time during the task, participants could view a specific rule on the computer screen by pressing the corresponding

number 1 through 7 on the keyboard. We operationally defined subtask performance for this action as the number of rules viewed during a trial.

Results

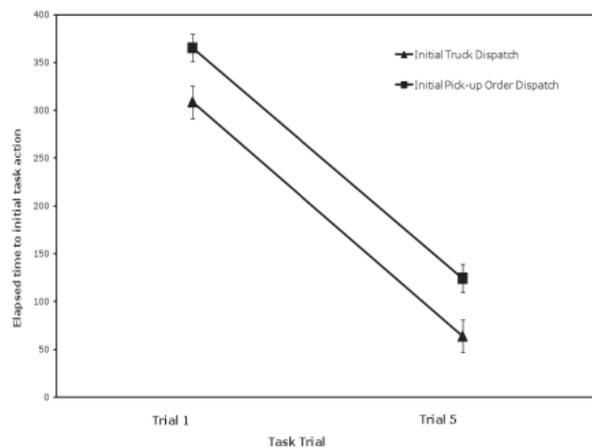
Task Action Speed as a Measure of Learning (Hypotheses 1 and 2)

To analyze the effects of task practice on participants' speed of task actions, we conducted a repeated measures analysis examining the effect of practice on time to complete initial task actions. Because we posited that practice would affect performance across all participants, regardless of learner type, we used the full sample ($N = 145$) in this analysis. Further, because points were awarded only when participants shipped units of office equipment, either on a truck or out through the pick-up window, two task actions essential to the successful completion of the task were "shipped truck" and "sent out pick-up."

If participants learned to more-efficiently complete the task with task practice, we would expect to see an effect of task practice (first vs. final trial) on both the time participants took to dispatch an initial truck and on the time participants took to dispatch an initial pick-up order shipment. In support of Hypothesis 1, there was a significant effect of trial on initial truck shipment speed, $F(1, 144) = 261.45$, $p < .0001$, *Wilks' Lambda* = .35, and on initial pick-up order shipment speed, $F(1, 144) = 153.90$, $p < .0001$, *Wilks' Lambda* = .48. The means for participants' speed of initial task actions in Trial 1 and Trial 5 are displayed in Figure 1. Participants performed initial task actions more rapidly in Trial 5

FIGURE 1

Mean elapsed time taken to complete two initial task actions for all participants during Trial 1 and Trial 5. Error bars indicate standard errors.



than in Trial 1.

We also expected that increased speed of initial truck shipment and pick-up order shipment would distinguish between those participants demonstrating the greatest learning and those demonstrating the least learning (Hypothesis 2). To test this prediction, we conducted a 2 x 2 mixed factorial analysis, with one between-subject variable (learner type) and one within-subject variable (trial). We used the reduced sample ($n = 97$) in this analysis, which included 49 participants in the greatest-learning group and 48 participants in the least-learning group.

Contrary to expectations, those in the greatest-learning group did not display significantly different improvements in their speed of initial truck shipment from the first to the final trial, compared to those in the least-learning group; performance improved similarly with task practice for both groups. Learner type had a significant main effect on speed of initial truck shipment, $F(1, 95) = 8.26, p = .01, Wilks' \Lambda = .38$. There was also a significant trial effect, $F(1, 95) = 153.46, p < .0001, Wilks' \Lambda = .38$, in the reduced sample, consistent with the results for the full sample used in testing Hypothesis 1. However, the learner type by trial interaction effect on speed of initial truck shipment was not significant. The means for the greatest-learning and the least-learning groups' speed of initial truck shipment in Trial 1 and after task practice in Trial 5 are displayed in Figure 2.

In contrast, those participants in the greatest-learning group improved significantly more in speed of initial pick-up order shipment from the first to the final trial, relative to those in the least-learning group.

The learner by trial interaction effect on speed of initial pick-up order shipment was significant, $F(1, 95) = 7.70, p = .01, Wilks' \Lambda = .92$, providing partial support for Hypothesis 2. There was also a significant effect of learner type on speed of initial pick-up order shipment, $F(1, 95) = 29.54, p < .0001$, as well as for trial, $F(1, 95) = 100.90, p < .0001, Wilks' \Lambda = .48$. The means for the greatest-learning and the least-learning groups' speed of initial pick-up order shipment in Trial 1 and Trial 5 are displayed in Figure 3.

Information-seeking Behavior as an Indicator of Learning (Hypothesis 3)

To examine the relationship between information-seeking behavior and learning, we examined the relationship between total number of rule call-ups during Trial 1 and final performance scores in Trial 5 using the full ($N = 145$) sample. Results did not indicate a significant correlation between rule call-ups during Trial 1 and Trial 5 performance scores, providing no support for Hypothesis 3.

Errors as a Measure of Learning (Hypotheses 4, 5, and 6)

To examine the relationship between initial errors and learning (Hypothesis 4), we examined the relationship between rule violations (errors) during Trial 1 and final performance scores in Trial 5 using the full ($N = 145$) sample. Results indicated significant positive correlations for Rule 2, $r = .29, p < .01$; Rule 5, $r = .21, p = .01$; and Rule 7, $r = .47, p < .01$, providing partial support for Hypothesis 4.

To test the prediction that types of errors would

FIGURE 2

Mean elapsed time taken to dispatch initial truck by learner type during Trial 1 and Trial 5. Error bars indicate standard errors.

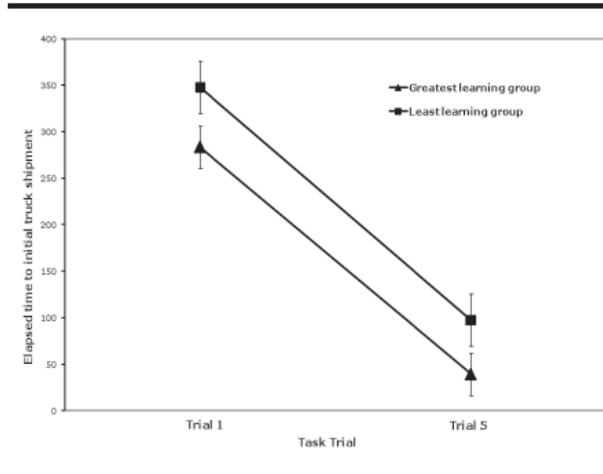
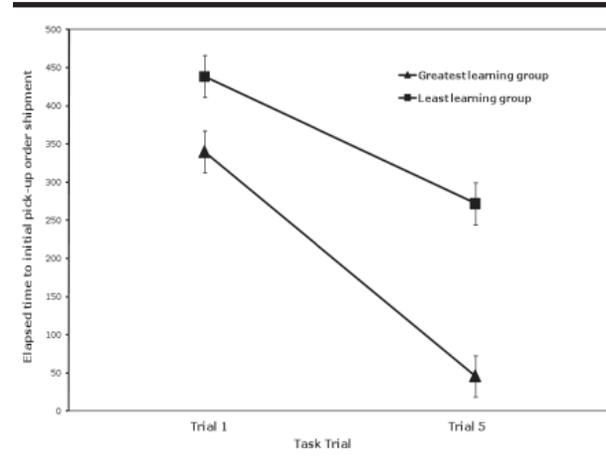


FIGURE 3

Mean elapsed time taken to dispatch initial pick-up order shipment by learner type during Trial 1 and Trial 5. Error bars indicate standard errors.



distinguish between those individuals in the greatest-learning group and those in the least-learning group (Hypothesis 5), we compared first the individual rule violations made by the participants in the greatest-learning group to those of the least-learning group in Trial 1, and then the individual rule violations made by the greatest-learning group to those of the least-learning group in Trial 5. Because our analyses included the learner variable, we used the reduced ($n = 97$) sample. For Trial 1, results indicated that participants in the greatest-learning group made significantly more violations relative to those in the least-learning group for Rule 1, $F(1, 95) = 5.15, p = .02$; Rule 2, $F(1, 95) = 6.90, p = .01$; and Rule 7, $F(1, 95) = 8.21, p = .01$. The means for the greatest-learning and the least learning groups' specific rule violations during Trial 1 are displayed in Figure 4. For Trial 5, results indicated that the greatest-learning group made significantly fewer violations than the least-learning group for Rule 1, $F(1, 95) = 36.90, p < .0001$; Rule 2, $F(1, 95) = 11.22, p < .01$; and Rule 7, $F(1, 95) = 77.29, p < .0001$. However, during Trial 5, the greatest-learners also made significantly more violations of Rule 3, $F(1, 95) = 8.69, p < .01$ and Rule 5, $F(1, 95) = 10.47, p < .01$. The means for the greatest-learning and the least-learning groups' specific rule violations during Trial 5 are displayed in Figure 5. In summary, results indicated that certain rule violations (types of errors) distinguished between participants who demonstrate the greatest learning and those who demonstrate the least learning, providing partial support for Hypothesis 5.

To examine our prediction that the greatest-learning group and the least-learning group would demonstrate different patterns of change in errors (Hypothesis 6), we conducted a 2 x 2 mixed factorial

analysis with one between-subjects factor (learner type) and one within-subjects factor (trial). We again used the reduced ($N = 97$) sample. We observed a significant trial by learner type interaction effect for Rule 1, $F(1, 95) = 38.02, p < .0001, Wilks' Lambda = .71$; Rule 2, $F(1, 95) = 19.84, p < .0001, Wilks' Lambda = .83$; Rule 3, $F(1, 95) = 8.87, p < .01, Wilks' Lambda = .91$; Rule 5, $F(1, 95) = 9.59, p < .01, Wilks' Lambda = .91$; and Rule 7, $F(1, 95) = 66.82, p < .0001, Wilks' Lambda = .59$, providing partial support for Hypothesis 6 in that changes in errors distinguished between participants demonstrating the greatest learning and those demonstrating the least learning.

Discussion

Our study had two main purposes. The first purpose was to determine if alternative measures of learning could be applied to training on our truck-dispatcher task. Our second purpose was to identify specific indicators of learning that occur during training, which could then be used to detect learning and to modify training parameters to improve training success. Using our entire data set, we evaluated the relationship between certain predictors (i.e., task speed, information seeking, and number of initial mistakes) and final performance. Using a reduced data set, we evaluated the effects of learner type on task speed and patterns of mistakes. Results indicated that both the speed with which participants took to complete an initial task action, such as shipping out an initial truck or dispatching an initial pick-up order, as well as the nature and pattern of mistakes, could be used as indicators of learning. Our results suggest that giving participants more specific information about these two dimensions of performance could have beneficial effects on training outcomes in similar tasks. Moreover, our results

FIGURE 4

Mean individual rule-violations by learner type during Trial 1. Error bars indicate standard errors.

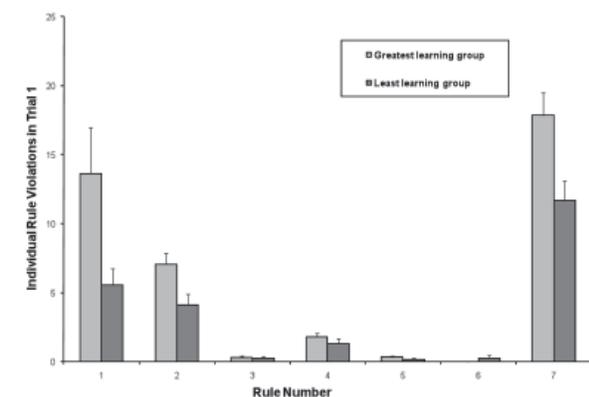
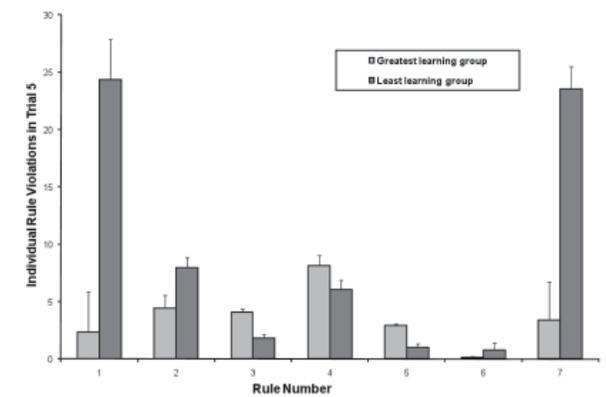


FIGURE 5

Mean individual rule-violations by learner type during Trial 5. Error bars indicate standard errors.



highlight the benefits of using multiple, alternative behavioral variables other than final performance scores as indicators of learning.

Task Speed as an Indicator of Learning

As expected, our results indicated that, overall, participants' speed of task actions improved with task practice for both dispatching initial trucks and pick-up orders. However, in complex tasks, some individuals may rapidly increase performance speed whereas others improve more slowly, sometimes not progressing beyond even a novice level (Ackerman & Beier, 2007). In the truck-dispatcher task, elapsed time taken to dispatch an initial pick-up order appeared to be an effective indicator variable for learning because it distinguished between participants who displayed the greatest amount of learning and those who displayed the least amount of learning. More specifically, we observed a significant trial by learner type interaction effect on time taken to dispatch an initial pick-up order. Those in the greatest-learning group became much faster at this skill by the final trial, relative to those in the least-learning group. This interaction effect was not observed for time taken to dispatch an initial truck. Dispatching trucks is possibly less challenging, as might be indicated by the shorter overall action completion times, whereas dispatching pick-up orders is more difficult (see Figure 1). Thus, speeding up pick-up order shipments could be a secondary strategy that further enhances performance after truck shipment is mastered. Future research should be done to further examine the effects of strategies on task understanding.

Information Seeking as an Indicator of Learning

Contrary to our hypothesis, our correlational analyses did not reveal a relationship between the amount of time participants spent re-reading task rules and final task performance. One possible explanation for this is that participants who focused on reading task rules during Trial 1 rather than immediately exploring the task did not gain as much actual task experience as others. Researchers (e.g., van der Linden et al., 2001) have noted that one type of behavior pattern that can arise when people are faced with a new and complex task is becoming too focused on obtaining more information, rather than exploring the task itself. Although having some information about a task is necessary for performance, if the new information leads to more uncertainty about how to approach the task and/or more information seeking, decisions and actions may be delayed or not taken at all, and participants could lose precious time to practice the task and/or learn from mistakes. Future research should determine when and how information-seeking behavior becomes

detrimental to performance.

Errors as Indicators of Learning

We did observe significant correlations between certain rule violations at the beginning of training (Trial 1) and performance scores at the end of training (Trial 5) using the full sample. This provided initial support for the hypothesis that trainees experiencing a greater number of errors early in training would gain more experience with the task and would demonstrate better final performance than those who did not experience these errors early on. These significant correlations led us to further examine participants' individual errors. However, we note that caution should be used when interpreting the correlations from the full sample because significant correlations are more likely in larger sample sizes.

Our analysis of errors indicated that those participants who displayed the most learning did not necessarily make fewer errors relative to those who displayed the least learning after task practice. Instead, different patterns of errors and different patterns of change in errors distinguished between the greatest- and the least-learning groups. In order to understand these differences in change and patterns, a thorough analysis of the individual errors is required. When comparing the two groups, we observed two interesting patterns of mistakes. One pattern was the reduction of certain errors (Rules 1, 2, and 7). The other pattern was the apparent use of errors (Rules 3 and 5) as learning opportunities beneficial to performance. A summary of the various rules and their implications may assist in understanding our results.

For Rules 1, 2, and 7, those in the greatest learning group originally made more violations relative to those in the least learning group during Trial 1; then they improved on these specific rules, making fewer violations of Rules 1, 2, and 7 during Trial 5 relative to those in the least learning group. Rule 1 indicated that participants must place units on a truck or send them to a pick-up order window based on the label associated with each unit. If participants placed an item meant for pick-up onto a truck or vice versa, the item would not ship out and the participant would lose points. Obeying this rule seems intuitively necessary for successful performance, especially because violating it does not allow delivery of units. The necessity to performance was demonstrated in that those individuals whose performance scores increased the most demonstrated a significant decrease in this error relative to those whose performance did not increase as much.

Rule 2 required participants to ship units labeled as "regular" within two task days (4 min in real time) and units labeled as "rush order" within one task day (2

min in real time). At the end of each working day, the number of “regular” and “rush orders” that remained in the queue rather than shipped out were counted and points were deducted. This rule is based on speed of performance. Those in the greatest learning group displayed a significant decrease in this error after task practice relative to those in the least learning group.

Rule 7 required participants to accept orders into their queues from the waiting list of orders within one task day (2 min in real time). Similar to Rule 2, this rule encouraged speed of task actions and distinguished between participants in the greatest-learning and the least-learning groups.

In contrast, those in the greatest-learning group actually violated Rules 3 and 5 more often than those in the least-learning group, indicating a potential strategic use of errors. Rule 3 required that participants dispatch trucks to no more than three out of five available shipment zones; and Rule 5 stated that participants were required to limit loads on each truck to that truck’s capacity. Whereas violating either of these rules resulted in points being subtracted from participants’ total score, ignoring these rules could enhance efficiency and total performance score. For example, if a participant dispatched a truck quickly to four instead of three zones or slightly overloaded a truck, he or she was able to deliver more units to customers, and his or her performance score benefited from this increase in productivity. Although the participant lost a few points for each rule violation, the lost points were more than compensated for by the rapid delivery of more fully-loaded trucks. These results are consistent with research examining performance quantity/quantity tradeoffs (see Austin & Bobko, 1985, for a discussion), research which has suggested that individuals might ignore or violate rules (i.e., make mistakes) if that allowed them to increase performance quantity.

Theoretical and Practical Implications

Learner type distinction. For the purposes of our analyses, we compared individuals who displayed the greatest increase in performance scores (top 33%) from early to final task practice to those individuals who displayed the least increase in performance scores (bottom 33%). This does not imply that those individuals who displayed the least increase had the lowest scores, in fact performance scores varied widely in both groups. We are also not ruling out the possibility that participants in the least-improvement group actually learned something during training as most participants demonstrated some improvement from beginning to end. However, we were interested in identifying patterns of behaviors indicative of a change in strategy use or task understanding, and those individuals who displayed the

least improvement in performance score did not demonstrate extreme changes in their behavior. Our results highlight the feasibility of identifying task features that are likely to distinguish learning rates. Such information might be useful in designing and delivering training interventions. For example, in the truck-dispatcher task, knowledge of trainees’ specific errors or speed of pick-up order shipment might enable trainers to adapt the training environment to enhance learning.

Using errors strategically. Our analyses demonstrated that for improved task success, participants forfeited certain rules in favor of increased production. This is certainly an example of the classic speed/accuracy trade-off (Austin & Bobko, 1985). As participants’ performance scores increased, they became faster at shipping trucks and dispatching pick-up orders, and they violated time-based rules (Rules 2 and 7) less often. However, although they managed to improve their scores and dispatch more units to customers overall, quality of performance declined. Trucks were slightly overloaded (Rule 5) or were dispatched to more than three zones (Rule 3). Ignoring these rules in favor of increased quantity of units delivered may be a strategy trainers could use to encourage increased performance speed when necessary. On the other hand, the ability to identify behaviors that indicate an individual is forfeiting quality for quantity may be beneficial for trainers who want to encourage both quality and quantity of performance for the long-term success of a company. Our results suggested that although task instructions did not emphasize quantity over quality or vice versa, participants made strategic decisions to sacrifice some quality to obtain gains in performance quantity. Trainers might be similarly able to strategically focus trainees on quantity, quality, or both, to obtain organizational goals.

Limitations

As with any study, there are limitations that should be addressed. First, the use of undergraduate college students as participants may make the ability to generalize findings to the intended population more difficult. College students may be less motivated to complete the experimental task than employees given training for their actual jobs. Due to time constraints, college students are also more likely to hold part-time rather than full-time jobs. However, our results should support generalizations to entry-level, college-educated employees in part-time positions. Additionally, our study included a much higher number of females than males, so generalizations based on our research to male-dominated professions should be made with caution.

Additionally, we were only able to obtain limited information about participants’ behaviors in the

current study. First, the variables recorded in the task simulation limited our ability to identify specific indicators of learning. The task simulation software recorded information included in performance score calculation (e.g., time at which trucks or pick-up orders were shipped or rules violated as well as the number of these events) but did not include all intermediate steps (e.g., moving items onto or off of a truck or from one truck to another). Knowledge of these intermediate steps may have improved our ability to identify specific indicators of learning. Also, we were only able to use scores and information generated during the first and final training trials to which participants were exposed. Although this may have limited our ability to determine exactly when trainees were learning during the task, it did allow us to observe specific changes in behavior from initial to final task practice.

Another limitation, which also relates to future research possibilities, is our operational definition of learner type. There are many possible ways to measure learning, and for the purposes of this study, we compared those participants whose performance scores displayed the greatest increase to those whose performance scores displayed the least increase. This comparison of the “most-improved” participants to those who did not display such changes in performance highlights the need for researchers and practitioners to specifically define a final criterion when evaluating training outcomes. Should training be considered more effective if participants demonstrate high scores or much improvement in scores? Certain people are top performers from the beginning of training and display little change in performance, whereas others show dramatic improvement from start to finish. Researchers could potentially assert that a high final performance score reflects the greatest learning during training; alternatively, examining improvement over the course of training is a separate area of interest in training effectiveness evaluation (e.g., Mathieu, Tannenbaum & Salas, 1992). Indeed, Sackett and Mullen (1995) suggested that a focus on a minimum criterion level of performance might be more useful for trainers than the typical focus on a criterion based on change or improvement.

Conclusion

In conclusion, we tested the validity of using alternative behavioral variables other than final performance scores as indicators of learning. Specifically, results indicated that for the truck-dispatcher task, task action speed and patterns of errors could be used as learning-indicator variables. Our results contribute to the literature by providing evidence that researchers can identify multiple behavioral indicators of learn-

ing, based on both speed and accuracy, and these can be applied to training improvement and research. We also contributed to the literature by providing evidence that people sometimes use errors strategically. Finally, our study highlights the notion that researchers should carefully consider whether to base assessments of training on final performance scores or improvement in performance across training and that their decisions should reflect the goals of the research and/or training they are conducting. By using an interdisciplinary approach to identify alternative measures of learning, this study opens up new avenues for future researchers who could use similar strategies to identify indicators of learning for other tasks.

References

- Ackerman, P. L., & Beier, E. (2007). Further explorations of perceptual speed abilities in the context of assessment methods, cognitive abilities, and individual differences during skill acquisition. *Journal of Experimental Psychology: Applied*, *13*, 249-272.
- Arthur, W., Bennett, W., Edens, P. S., & Bell, S. T. (2003). Effectiveness of training in organizations: A meta-analysis of design and evaluation features. *Journal of Applied Psychology*, *88*, 234-245.
- Ashford, S. J., & Cummings, L. (1983). Feedback as an individual resource: Personal strategies of creating information. *Organizational Behavior & Human Performance*, *32*, 370-398.
- Austin, J. T., & Bobko, P. (1985). Goal-setting theory: Unexplored areas and future research needs. *Journal of Occupational Psychology*, *58*, 289-308.
- Baldwin, T. T., & Ford, K. (1998). Transfer of training: A review and directions for future research. *Personnel Psychology*, *41*, 63-105.
- Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Englewood Cliffs, NJ: Prentice Hall.
- Berninger, V. W., Abbott, R. D., Jones, J., Wolf, B. J., Gould, L., Anderson-Youngstrom, M., et al. (2006). Early development of language by hand: Composing, reading, listening, and speaking connections; Three letter-writing modes; and fast mapping in spelling. *Developmental Neuropsychology*, *29*, 61-92.
- Cascio, W. F., & Aguinis, H. (2005). *Applied psychology in human resource management*, (6th Ed.). New Jersey: Prentice Hall.
- Dormann, T., & Frese, M. (1994). Error training: Replication and the function of exploratory behavior. *International Journal of Human-Computer Interaction*, *6*, 365-372.
- Erez, M., (1990). Performance quality and work motivation. In V. Kleinbeck, H. Thierry, H. Haecker & H. Quast (Eds.), *Work motivation* (pp. 53-65). Hillsdale, NJ: Erlbaum.
- Erez, M., & Arad, R. (1986). Participative goal-setting: Social, motivational, and cognitive factors. *Journal of Applied Psychology*, *71*, 591-597.
- Facteau, J. D., Dobbins, G. H., Russell, J. E. A., Ladd, R. T., & Kudisch, J. D. (1995). The influence of general perceptions of the training environment on pretraining motivation and perceived training transfer. *Journal of Management*, *21*, 1-25.
- Gettlinger, M. C., & White, M. A. (1979). Which is the stronger correlate of school learning: Time to learn or measured intelligence? *Journal of Educational Psychology*, *71*, 405-412.
- Hunter, J. E., & Schmidt, L. (1996). Intelligence and job performance: Economic and social implications. *Psychology, Public Policy, and Law*, *2*, 447-472.
- Ilgen, D. R., Fisher, C. D., & Taylor, M. S. (1979). Consequences of individual feedback on behavior in organizations. *Journal of Applied Psychology*, *64*, 349-371.
- Ivkovich, D., Eckerman, C. O., Krasnegor, N. A., & Stanton, M. E. (2000). Using eyeblink conditioning to assess neurocognitive development. In D. S. Woodruff-Pak & J. E. Steinmetz (Eds.), *Eyeblink classical conditioning, volume I: Applications in humans* (pp. 119-142). Amsterdam: Kluwer Academic Publishers.

- Keith, N., & Frese, M. (2008). Effectiveness of error management training: A meta-analysis. *Journal of Applied Psychology, 93*, 59-69.
- Kraiger, K., Ford, J. K., & Salas, E. (1993). Application of cognitive, skill-based, and affective theories of learning outcomes to new methods of training evaluation. *Journal of Applied Psychology, 78*, 311-328.
- Kraiger, K., & Jung, K. M. (1997). Linking training objectives to evaluation criteria. In M. Quinones & A. Ehrenstein (Eds.), *Training for a rapidly changing workplace: Applications of psychological research* (pp 151-176). Washington, DC: American Psychological Association.
- Lorenzet, S. J., Salas, E., & Tannenbaum, S. I. (2005). Benefiting from mistakes: The impact of guided errors on learning, performance, and self-efficacy. *Human Resource Development Quarterly, 16*, 301-322.
- Mathieu, J. E., Tannenbaum, S. I., & Salas, E. (1992). Influences of individual and situational characteristics on measures of training effectiveness. *Academy of Management Journal, 35*, 828-847.
- Palumbo, M. V. (2008). Cognitive ability, job knowledge, and stereotype threat: When does adverse impact result? (Doctoral dissertation, Wright State University, 2007). *Dissertation Abstracts International, 68*, 140.
- Quiñones, M. A. (1997). Contextual influences: on training effectiveness. In M. Quinones & A. Ehrenstein (Eds.), *Training for a rapidly changing workplace: Applications of psychological research* (pp 177-199). Washington, DC: American Psychological Association.
- Sackett, P. R., & Mullen, E. J. (1995). Beyond formal experimental design: Towards an expanded view of the training evaluation process. *Personnel Psychology, 46*, 613-628.
- Schmidt, F. L., & Hunter, E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin, 124*, 262-274.
- Singer, R. N., & Gaines, L. (1975). Effects of prompted and problem-solving approaches on learning and transfer of motor skills. *American Educational Research Journal, 12*, 395-403.
- Skinner, B. F. (1953). *Science and human behavior*. New York: Free Press.
- Stanton, J. M., & Julian, L. (2002). The impact of electronic monitoring on quality and quantity of performance. *Computers in Human Behavior, 18*, 85-101.
- Steele-Johnson, D., & Hyde, B. G. (1997). Advanced technologies in training: Intelligent tutoring systems and virtual reality. In M. Quinones & A. Ehrenstein (Eds.), *Training for a rapidly changing workplace: Applications of psychological research* (pp 225-248). Washington, DC: American Psychological Association.
- Steele-Johnson, D., & Perlow, R. (1989). *Computer-generated truck-dispatch simulation task* [Computer software]. Houston, TX: University of Houston.
- van der Linden, D., Sonnentag, S., Frese, M., & van Dyck, C. (2001). Exploration strategies, performance, and error consequences when learning a complex computer task. *Behaviour & Information Technology, 20*, 189-198.
- Warr, P., & Bunce, D. (1995). Trainee characteristics and the outcomes of open learning. *Personnel Psychology, 48*, 347-375.

APPENDIX

Task Rules

- Rule 1.** Participants had to correctly identify whether units were to be placed on a truck, or sent to the pick-up order window based on the label associated with the unit.
- Rule 2.** Participants were required to ship units labeled as "regular" within two task-days (4 min in real time), and units labeled as "rush order" within one task-day (2 min in real time).
- Rule 3.** Participants were required to dispatch trucks to no more than three out of five available shipment zones.
- Rule 4.** Each of the units displayed were labeled with a particular business name. Units with the same business names were to be shipped together on one truck rather than dispersed between several trucks.
- Rule 5.** Each truck had a limited capacity, and participants were required to limit loads on each truck to that capacity.
- Rule 6.** Trucks exceeding their capacity by more than 20% were not dispatched.
- Rule 7.** Orders first appeared in the "queue," which was a waiting list of orders not yet accepted. Participants were required to accept orders from the queue within one task day (2 min in real time).
-

College Students' Perceptions of Alcohol Use in Society and in Entertainment Television

LINDSAY G. COOK

RICHARD JACKSON HARRIS*

Kansas State University

The present study explored college students' use of alcohol and their attitudes toward alcohol abuse in society. The study also examined the incorporation of Entertainment-Education concepts within entertainment television by looking at how specific programs structure and emphasize information on alcohol use. Participants who typically drank more than 7 drinks in one sitting (N = 47) rated alcohol abuse in society as significantly less serious than participants who typically drank fewer than 7 drinks in one sitting (N = 139). They were also significantly more entertained by programs that included alcohol use. Although 8 of the top 11 most-reported themes were prodrinking in nature, the results indicated that many programs also incorporated prohealth messages pertaining to alcohol use.

Despite the general consensus that consuming large quantities of alcohol can be dangerous, research has shown that alcohol abuse by college students continues to be an issue (Hingson, Heeren, Zakocs, Kopstein, & Wechsler, 2002; Knight et al., 2002; National Institute of Alcohol Abuse and Alcoholism [NIAAA], 2004; Wechsler, Lee, Kuo, & Lee, 2000). Approximately 80% of college students drink and 23% of undergraduates binge drink more than once a week (NIAAA, 2004; Wechsler et al., 2000). Binge drinking is defined as drinking that leads to a blood alcohol concentration (BAC) of .08% or more, which is achieved through the consumption of approximately 4-5 alcoholic beverages within a 2-hour time span (NIAAA, 2004). According to criteria outlined in the DSM-IV, 31% of undergraduates are classified as alcohol abusive (Knight et al., 2002). Finally, 2.1 million students, out of a sample of 8 million students, indicated they had driven while intoxicated (Hingson et al., 2002). Three million of these individuals indicated they had ridden with a driver who they knew to be intoxicated.

Many media campaigns directed toward young people and their drinking behaviors have been created in response to these and similar statistics. Petty, Priester, and Brinol (2002) outlined the two components of a successful campaign. First, the message must be effective in appropriately altering the recipients' attitudes. Second, these altered attitudes must cause recipients to make changes in their corresponding behaviors.

In the past—and still today—many variations of public service announcements (PSAs) were incorporated into the media in hopes of altering unhealthy or dangerous behaviors (Brown & Walsh-Childers, 2002; Harris, 2004). Unfortunately, however, there are many limitations to this social marketing technique (Harris, 2004). For example, PSAs on alcohol use are subject to limited funding, appear less frequently on television, and are generally of poorer quality than prodrinking images and advertisements. PSAs are also rarely aired during peak television-viewing hours, which makes PSAs less likely to reach the target audience (Brown & Walsh-Childers, 2002; Harris, 2004).

Acknowledging the limitations of the PSA strategy, more promising techniques have been implemented to create change through media—television in particular. Considering that the average person will watch 3 to 4 hours of television every day, television programs are an attractive medium for social marketing campaigns (Brown & Walsh-Childers, 2002). Beyond the amount of television being viewed, the media in general are perceived as a likely source for social marketing campaigns. Previous research has shown that the media are able to effectively provide information about and models for health-related attitude and behavior changes on an individual level (Brown & Walsh-Childers, 2002).

Currently, especially in developing, third-world

* Faculty mentor

countries, the technique of Entertainment-Education (E-E) has been used to replace or supplement PSAs. E-E incorporates educational messages—designed to inform and influence viewers' attitudes and behaviors pertaining to their health—into an entertainment format (Kennedy, O'Leary, Beck, Pollard, & Simpson, 2004). Characters portrayed in these programs are rewarded for making recommended behavioral changes related to their health, such as depicting a promising future after taking necessary steps to get sober. Other characters are severely punished for not adopting the proposed changes, such as dying in a car wreck due to drunk driving. Finally, a third group—called transitional characters—are included to inspire audience identification. These characters observe the health issue from many perspectives, decide to alter their own attitudes and behaviors despite overwhelming obstacles, succeed in making positive changes, and are finally rewarded in a way that inspires the target audience to take action (Kennedy et al., 2004).

Theoretical support for E-E comes from Bandura's social cognitive theory (Bandura, 1986, 2002). Bandura's theory states that modeling a particular behavior as well as the consequences can effectively induce learning in the viewer (Kennedy et al., 2004; Trull, 2005). It is also important that the role model encounter and overcome likely barriers to reaching an overarching goal. The model must build the viewer's self-efficacy in addition to supporting the notion that attitude and behavior changes will result in a positive outcome (Bandura, 1986, 2002). Thus, Bandura's social cognitive theory suggests that an E-E television program has value in that it can encourage persistence in the face of barriers and depict positive consequences as a result of adopting the modeled attitude or behavior change (Brown & Walsh-Childers, 2002).

In addition to Bandura's social cognitive theory, the health belief model lends theoretical support to E-E (Rosenstock, 1974). This theory posits that health behavior is influenced by (a) how susceptible individuals feel they are to a disease or outcome, (b) the perceived seriousness of a disease or behavior, (c) how great individuals perceive benefits of behavior change to be, (d) the lack of overwhelming barriers to taking action, and finally (e) cues to action. In this way, serialized drama, such as a weekly television program, provides a format in which the prevalence and severity of a disease or behavior may be introduced, followed by an illustration of the benefits of taking action, and finally suggestions on how to avoid or alter dangerous behaviors.

Along with theoretical underpinnings, there is also practical support for the efficacy of E-E programs. First, studies have shown that information has a greater

impact on the audience when presented within a novel storyline, such as within an E-E program, rather than through strictly informative material (Brinson & Brown, 1997; Parrott, 1995). Second, the nature of serialized television allows the storyline of previous episodes to be repeated and developed in future episodes. This characteristic is important according to media theory, which states that repetition of a theme is important in increasing awareness of an issue (Rosenzweig, 1999; Sherry, 2002). In addition, when an individual's favorite character on a popular television show discusses a health-related issue within the context of the program, the viewer is likely to attend to the information, which provides the opportunity for learning—and later attitudinal or behavioral change (Papa et al., 2000; Sherry, 2002). More specifically, because most individuals view the same shows every week, parasocial relationships (PSRs) and parasocial interactions (PSIs) are likely to develop and perhaps trigger referential involvement—or, the consideration of behavior change as a result of how well the viewer is able to relate to the message (Papa et al., 2000). Klimmt, Hartman, and Schramm (2006) identified PSRs as “a special type of ‘interpersonal involvement’ that combines different phenomena such as interaction, identification, and long-term identification with media personae” (p. 292). PSIs are those in which viewers may talk at the character (i.e., speak to the television as though the character can hear them), invest cognitive effort thinking about the fictional character, experience similar emotions as the character, or find the mood of the fictional character contagious (Klimmt et al., 2006; Papa et al., 2000). PSIs may also prompt viewers to engage in more postviewing cognitions and discussions with others about the characters or the show—thus creating a social learning environment—in contrast to random television watching (Klimmt et al., 2006; Papa et al., 2000). Due to the development of these relationships and interactions, the characters are highly effective in inspiring active message processing, which improves memory for the information and thus is more likely to influence behavior change (Kennedy et al., 2004; Larson, 1991; Papa et al., 2000; Parrott, 1995).

As indicated, in recent years producers have begun incorporating E-E into their programming. For example, in many developing countries, radio and television programs have been written specifically to convey prohealth messages and information (Harris, 2004). From 1975-1982 Mexico's private network, Televisa, aired seven telenovelas (soap operas) that promoted gender equality, adult literacy, sexual responsibility, and family planning (Brown & Singhal, 1990; Rogers & Singhal, 1990). These popular shows were effective in that viewers often requested services discussed within

the programs (Lozano, 1992).

With Televisa's success in promoting prosocial and prohealth messages, many other countries were inspired to try similar methods. In 1987, Kenya began broadcasting the soap opera *Tushauriane* ("Let's Discuss") that focused on family planning. The show became Kenya TV's most popular program to date (Brown & Singhal, 1990). From 1993-1997, the Tanzanian radio soap opera *Twende na Wakati* ("Let's Go With the Times") had measurable effects on the population's adoption of family planning methods. In this soap opera, Tunu, the wife of an alcoholic truck driver, adopts a family planning method and then leaves her HIV-positive husband. Tunu is rewarded for her decisions with a better life for her family and herself. Her actions as a transitional character were especially important for listeners because they modeled self-efficacious behavior, which is important for viewers/listeners in their decision to take action (Rogers et al., 1999).

Another example of successful E-E programming can be seen in South Africa's TV, radio, and public health campaign, *Soul City*, which dealt with HIV and tuberculosis prevention and control, maternal and child health, housing and urban reform, domestic violence, violence against women, youth sexuality, hypertension, and alcohol and tobacco abuse (Singhal & Rogers, 1999). The enormous success of the campaign is attributable to its multimedia approach: a prime-time TV drama, a radio drama series broadcast daily in eight different languages, health-education booklets, newspaper coverage, and community recognition and events. One example of the campaign's success is that, before *Soul City*, only 3% of the population indicated that it was necessary to tell one's partner if one is HIV positive. Not long after *Soul City* began, however, 75% of the population indicated that it was necessary to tell their partner if they had the disease (Harris, 2004).

Despite the recognized success of E-E programs in other countries, there is some concern about its potential for success in more media-saturated areas such as the United States (Sherry, 2002). Specifically, the prevalence of prodrinking messages makes it difficult for viewers to accept the prohealth messages. Also, private ownership of the media as well as First Amendment concerns have made it difficult for the government and nonprofit organizations to remove negative or unhealthy yet popular messages—such as consuming large quantities of alcohol—in order to deliver lengthy prosocial messages through entertainment media (Brown & Walsh-Childers, 2002).

To date, no long-running E-E campaigns have been created for entertainment television in the United States. Instead, some producers have agreed to collaborate with public health professionals, including

the CDC, to incorporate elements of E-E into their entertainment programs (Kennedy et al., 2004). For example, in the 1980s, Professor Jay Winsten of the Harvard School of Public Health worked with writers, producers, and executives to incorporate into television a new concept: the use of a designated driver (Rosenzweig, 1999). It is likely that this concept contributed to the significant decrease in drunk-driving fatalities recorded by the late 1990s. Other examples of E-E programming include AIDS awareness messages appearing on popular soap operas such as *General Hospital*, procondom messages on *Friends*, and a rape hotline appearing on the screen following a two-part episode involving rape on the show *Felicity* (Brown & Walsh-Childers, 2002).

More recently, research has been done to study the direct impact of E-E messages. In August 2001, a research group monitored calls to an AIDS hotline following two episodes of the popular soap opera, *The Bold and The Beautiful* (Kennedy et al., 2004). In the first episode, one of the characters was diagnosed with HIV. In the second episode, the HIV-positive character told his fiancée about his disease. Following both episodes, the number of phone calls to the provided hotline rose dramatically. Of the callers, 57% indicated that they intended to make a change or take action after seeing the show, 44% said they intended to be tested, and 28% said they intended to use a condom.

Research on the use of E-E in television programming has made it possible to identify specific factors that influence the overall effectiveness of a program's message (Rogers & Singhal, 1990). As discussed previously, drawing out a novel storyline over several episodes, creating memorable and appealing characters with whom viewers are able to create PSRs or to have PSIs, and providing suggestions for future actions are very important in E-E (Brinson & Brown, 1997; Kennedy et al., 2004; Klimmt et al., 2006; Papa et al., 2000; Parrott, 1995). In addition to these techniques, the producer and researchers should carefully consider their target audience in order to tailor the message appropriately as well as depict small steps in behavior change rather than suggesting a total attitude or behavior overhaul (Harris, 2004). In general, emotional appeals have been found to be more effective than cognitive appeals in stimulating someone to learn about a health-related issue. Along the same lines, it appears that the message should not be too blatant. PSAs have been criticized for trying to "sell" a behavior, which actually causes resistance in the viewers (Brown & Walsh-Childers, 2002).

Present Study

In response to the statistics on college students' abuse of alcohol as well as the average person's rate of televi-

sion consumption per day, the present study explored various portrayals of alcohol use and abuse within programs watched by college students. In contrast to the restrictions placed on cigarette advertising, alcohol ads—especially for beer and wine—appear frequently in the media (Brown & Walsh-Childers, 2002). For example, prodrinking references outnumber negative ones by more than a 10 to 1 margin on television and in movies (Brown & Walsh-Childers, 2002). These messages do not go unnoticed. Studies show that there is a direct relationship between alcohol advertising and positive associations with alcohol and intentions to drink (Brown & Walsh-Childers, 2002).

Therefore, it is necessary to explore what messages are being delivered to college students through entertainment television programs as well as what impact the information has on these viewers. The overall purpose of the present study was twofold: (1) to explore what information pertaining to the use of alcohol (prodrinking versus prohealth or E-E) is being portrayed in current entertainment television watched by college students, and (2) how college students perceive this information. In the first section of the survey, we asked questions pertaining to participants' current use of alcohol in addition to how they perceive alcohol use in society. Participants then used their autobiographical memory to answer questions about a television show that had portrayed one or more characters abusing alcohol.

We hypothesized that approximately 80% of our sample would drink, which is based on the findings from previous studies concerning college students' drinking habits (NIAAA, 2004; Wechsler et al., 2000). Second, we hypothesized that nearly all participants would be able to recall an entertainment television show that had portrayed drinking as a health-related issue. This hypothesis was based on research that shows the average person consumes 3 to 4 hours of television on a daily basis coupled with the knowledge that during the average hour of prime-time television, there will be several alcohol references (Brown & Walsh-Childers, 2002). Also, prodrinking and prohealth messages appear during both peak and nonpeak viewing hours, which indicates that any viewer could be subject to references to alcohol (Brown & Walsh-Childers, 2002; Harris, 2004). Next, we hypothesized that the participants who do consume alcohol will be more entertained by the programs and the episodes will hold their attention better than those who do not drink any alcohol. This hypothesis is based on the concept that the drinking group will be able to more easily identify with the characters who drink on the programs and/or the events surrounding the characters' use of alcohol. Finally, we hypothesized that although the majority of themes

pertaining to alcohol use within the programs watched by college students would be prodrinking in nature, we also expected some incorporation of E-E themes.

Method

Participants

The sample for this exploratory study was composed of 189 undergraduates (100 men, 89 women) enrolled in general psychology at a large midwestern university. The age range was 18 to 25 years with a mean age of 19. Finally, 90% of students were Caucasian, 3% were African American, 3% were Asian, 3% were Hispanic, and .5% were Alaskan.

Materials

For the study, the authors constructed a two-part exploratory survey (see Appendix for complete survey). The first section focused on the participants' demographic information, drinking behaviors and patterns, their perception of alcohol use in society, and whether they knew anyone who abuses or has abused alcohol. The number of drinks that participants reported they typically consume in one sitting determined drinking behavior. We determined four drinking level groups by considering NIAAA's definition of binge drinking, or approximately 4-5 alcoholic beverages in a two hour time span (NIAAA, 2004). Therefore, the first group included participants who reported that they typically drink less than what is considered to be binge drinking, or 1-3 drinks at one time. The next group included participants who reported that they consume approximately the minimum level of consumption necessary to be defined as binge drinking, or 4-6 drinks at one time. The third group consisted of participants who reported that they typically consume above the approximate minimum level of consumption necessary to meet the definition of binge drinking, or 7 or more drinks in one sitting. The final group consisted of participants who reported that they do not drink alcohol.

Section two of the survey focused on the participants' autobiographical memory for a particular entertainment television show of their choosing, which had "portrayed alcohol abuse as a health-related issue for one or more of the characters." It is important to note that although autobiographical memory may be flawed, the purpose of the present study was to understand how college students perceived and remembered the information they viewed. Thus, perception was more important to this study than accuracy of memory.

The authors constructed the second section of the survey by taking into consideration previously-defined, successful E-E concepts. The purpose of many of these questions was to discover what messages pertaining to alcohol use are incorporated in television programs

watched by college students and whether or not these messages are incorporating E-E methods. For example, the question about the number of episodes dedicated to the topic is relevant because repetition and development of a theme is an important component of E-E (Rosenzweig, 1999; Sherry, 2002). This question revealed how many programs are using this tool. Other examples would be the questions about how often viewers watched the show, how entertaining viewers found the episodes to be, and how well the episodes kept the viewers' attention. Again, these questions are important because the results could show effective use of E-E methods. As discussed, using a novel storyline as well as compelling characters with whom viewers can create PSRs is important in the efficacy of E-E (Brinson & Brown, 1997; Papa et al., 2000; Parrott, 1995). If viewers watch the show regularly, are entertained by it, and find that the episodes are able to keep their attention, this is perhaps an indication that the viewers are creating PSRs with the characters on screen as well as, generally, attending to the information.

On the second section of the survey, participants identified what alcohol-related themes were included in their television program. This section of the survey also asked participants about how the nonalcohol-abusing characters were impacted. For both questions, participants viewed a list and selected all themes that applied to their program. The researchers created these two lists by compiling an inventory of the themes pertaining to alcohol use and abuse, as well as how nonalcohol-abusing individuals are impacted by the drinking of those close to them, which were mentioned within the literature reviewed for this study. Both lists of themes were created to directly address the question of whether or not E-E themes are incorporated into the programs watched by college students.

Procedure

This survey was included in a packet of unrelated questionnaires administered to students during a mass testing session held after class. A graduate student from the department hosted the session. Students attended the mass testing in order to earn credit toward their required number of research participation hours for general psychology. Any student enrolled in general psychology was invited to attend the mass testing session.

Results

Of the 186 participants, 18% indicated that they did not drink ($N = 33$), 31% responded that they typically drank 1 to 3 drinks in a sitting ($N = 57$), 26% drank 4 to 6 drinks in one sitting ($N = 49$), and 25% drank 7 or more drinks in one sitting ($N = 47$). These results are consistent with our first hypothesis, which was based

on previous research showing that approximately 20% of college students do not drink (NIAAA, 2004; Wechsler et al., 2000). Subjects' drinking behavior became the first quasi-experimental variable used for analysis. Specific questions then became the isolated dependent variables.

The first isolated dependent variable considered participants' views on the severity of alcohol abuse as an issue in society, which is an important factor in E-E according to the health belief model (Rosenstock, 1974). In the first section of the survey, the following question was asked: "Do you feel alcohol abuse is an issue in our society?" These results are provided in Table 1. Using drinking behavior as the independent variable, statistical significance between the means was revealed in a one-way ANOVA, $F(3, 182) = 5.34, p < .05$, with simple main effects tests showing that the heaviest-drinking group rated alcohol abuse as a significantly less serious issue than any of the other three groups on a scale from 1 (*not at all an issue*) to 7 (*very severe issue*). There was a general trend indicating that the more a person drank in one sitting, the less apt they were to see alcohol abuse in society as an issue.

To the question "Do you know anyone personally who abuses or has abused alcohol?" 72% of participants ($n = 185$) answered yes, 16% answered no, and 12% indicated that they were unsure. If participants answered yes to this question, they were then asked to answer the question "How could you tell they abused alcohol?" All of the participants' written explanations of alcohol abuse met the definition criteria for alcohol abuse and/or alcohol dependence as outlined by the American Psychiatric Association (*DSM-IV-TR*, 2000).

Of the 189 participants, 77% could recall a television show that fit the survey criteria as outlined at the beginning of section two of the survey. This finding

TABLE 1

| Mean Responses on a 7-Point Scale to the Question "Do You Feel Alcohol Abuse is an Issue in Our Society?" | | |
|---|-------------------|------------|
| Drinks at one time | <i>M</i> | <i>SEM</i> |
| None ($n = 33$) | 5.42 | 0.24 |
| 1-3 ($n = 57$) | 5.49 ^a | 0.16 |
| 4-6 ($n = 49$) | 5.02 ^a | 0.18 |
| 7 or more ($n = 47$) | 4.55 ^b | 0.19 |

^aThese two groups are different from one another at $p < .05$

^bDifferent from all other groups at $p < .05$

is consistent with our second hypothesis that nearly all participants would be able to recall an entertainment television show that had portrayed drinking as a health-related issue. These participants listed 61 different television programs in the following categories: 40% were comedy/sitcom, 36% were drama, 18% were reality television, and 3% were soap operas. Also, approximately 78% of these participants (i.e., 112), reported that they watch the show *at least occasionally to almost always*. Only 22% reported that they *did not*, or *usually did not*, watch the show regularly that they had indicated on the survey. According to responses, approximately 50% of the shows dedicated three or fewer episodes to the topic of alcohol abuse, 27% dedicated 4-9 episodes to the topic, 14% dedicated 10 or more episodes to the topic, and 12% dedicated an entire season to alcohol abuse. Also, 55% of the alcohol-abusing characters listed were men, 32% were women, and 10% of the shows had both male and female alcohol-abusing characters. Finally, the range for the characters' ages—according to participants—was 16 to 60 years ($M = 31$), with participants reporting that 18% of the alcohol-abusing characters were under the legal drinking age of 21.

Participants then rated how entertaining the episodes were on a scale from 1 (*not at all entertaining*) to 7 (*very entertaining*) as well as the episodes' ability to keep the viewer's attention on a scale from 1 (*not very well*) to 7 (*very well*). The overall mean for how entertaining the viewers found the episode(s) of their show to be was 4.64. The overall mean for how well the episodes kept the viewers' attention was 5.04.

The authors used the perceived entertainment value of the episodes as a dependent variable relative to subjects' drinking behavior. A one-way ANOVA, followed by simple main effects tests, revealed statistical

significance between the mean rating of perceived entertainment value of the episodes by the group of participants who drank 7 or more drinks in one sitting, and the mean entertainment ratings of the three other groups, $F(3, 139) = 3.44, p < .05$ (see Table 2). These results confirmed the first part of our third hypothesis that the participants who do consume alcohol will be more entertained by the programs than those who do not drink any alcohol, but was more specific than predicted. Persons who drank more than 7 drinks in one sitting found the episodes to be significantly more entertaining than all of the other groups. Also, it is important to know that regardless of their drinking group, participants watched programs with similar pro-drinking and prohealth themes portrayed in them.

The episodes' ability to keep the viewer's attention also differed as a function of how much participants drank (see Table 2). As predicted, a one-way ANOVA, $F(3, 139) = 4.55, p < .05$, which was followed by simple main effects tests, revealed a significant difference between the mean rating of the episodes' ability to keep the viewers' attention by the group who reported they do not drink, and the other three groups. Specifically, the show was significantly less likely to keep the attention of those individuals who reported they do not drink. This finding confirms the second part of our third hypothesis that those who consume alcohol will report that the episodes kept their attention better than those who do not drink any alcohol.

A second quasi-independent variable examined in the first section of the survey was whether the participant knew someone who abuses or abused alcohol. The participants who knew someone who abuses or abused alcohol gave a mean rating of 5.29 on the scale inquiring about how well the episode(s) kept their attention. The individuals who did not know anyone who abuses

TABLE 2

Mean Ratings of Shows' Entertainment Value and Ability to Keep Viewers' Attention

| Drinks at one time | Entertainment | SEM | Attention | SEM |
|--------------------|-------------------|------|-------------------|------|
| None | 3.90 | 0.38 | 4.00 ^b | 0.38 |
| 1-3 | 4.43 | 0.26 | 4.92 | 0.26 |
| 4-6 | 4.58 | 0.28 | 5.12 | 0.28 |
| 7 or more | 5.28 ^a | 0.27 | 5.62 | 0.23 |

^aDifferent from all other groups at $p < .05$

^bDifferent from all other groups at $p < .05$

or abused alcohol gave a mean rating of 4.41 on the attention scale. Finally, the participants who were unsure whether or not they knew someone who abuses or abused alcohol gave a mean rating of 4.27. Statistical significance between the means was found using a one-way ANOVA, $F(2, 141) = 4.34, p < .05$, which indicated that the attention rating was significantly higher for those participants who knew someone who abuses or abused alcohol than for those participants who did not know anyone who abuses or abused alcohol.

A fourth question on the second part of the survey asked "Would you say the person (i.e. the character on the show) abusing alcohol was an alcoholic?" This question was answered by 145 participants. Of those, 68% marked *yes* or *probably yes*, 20% marked *not sure*, and 12% marked *no* or *probably no*. According to participants' autobiographical memory ($n = 144$), only 32% of all characters received treatment (19% of participants were unsure if characters were treated) and only 13% were treated effectively according to participants' perceptions of the shows' events.

Next, the two most-often reported genres of television shows (comedies and dramas) were partitioned by whether the character who abused alcohol was treated (see Table 3). According to a chi-square test, $X^2(2, N = 109) = 12.46, p < .05$, there was a significant difference between the two types of shows. Participant responses indicated that only 17% of characters in the comedy shows were treated while 41% of the characters featured in the dramas received treatment.

All of the alcohol-related themes identified by participants as being included in their television program were further partitioned by show genre. Table 4 provides a list of the themes from the survey, which are organized in descending rank order based on the percentage of participants who indicated that a theme from the survey was included in their television show.

In addition to treatment, results indicated that comedies and dramas differed significantly in their portrayal of two of the themes listed (see Table 4). Using

one-way ANOVAs, dramas were found to be significantly more likely than comedies to provide suggestions for how to communicate about drinking, $F(1, 105) = 4.10, p < .05$, and significantly less likely to portray drinking as a fun recreational activity, $F(1, 105) = 5.75, p < .05$.

Similarly, participants identified how the nonalco-

TABLE 4

Percentages of Participants Who Reported Themes Present in Entertainment Programs

| Theme | Percentage Reporting Theme |
|--|----------------------------|
| Drinking to relieve stress | 70 |
| Drinking to forget | 61 |
| Drinking to avoid emotions | 58 |
| Drinking to loosen up | 56 |
| Drinking as a fun recreational activity | 55 ^a |
| Drinking as a normal activity | 54 |
| Emotional issues (result) | 50 |
| Drinking to avoid responsibility | 43 |
| Drinking to fit in socially | 42 |
| Change in perception of why people drink | 37 |
| Need for moderation | 37 |
| Physical health issues (result) | 36 |
| Unplanned sex (result) | 35 |
| Emotional support for abusers | 35 |
| Increase in awareness of alcohol abuse | 35 |
| Need for designated driver | 32 |
| Emotional support for family/friends | 30 |
| Change in perception of who drinks | 30 |
| Treatment options | 29 |
| Increase in knowledge about alcohol abuse | 29 |
| Change in attitude about alcohol abuse | 25 |
| Suggestions for communicating about drinking | 21 ^b |

^aComedies significantly more likely to portray than dramas at $p < .05$

^bDramas significantly more likely to portray than comedies at $p < .05$

TABLE 3

Percentages of Answers to the Question "Was This Character Treated for Alcohol Abuse?"

| Show Genre | Treated | | |
|----------------------------------|---------|----|----------|
| | Yes | No | Not Sure |
| Comedy ($n = 58$) ^a | 17 | 71 | 12 |
| Drama ($n = 51$) ^a | 41 | 37 | 22 |

^aComedies and dramas different from one another at $p < .05$

hol-abusing characters on the programs were impacted by selecting themes from a list. Table 5 provides the list of themes from the survey as well as the percentage of participants who selected each theme. The authors partitioned the list of themes pertaining to the impact on nonalcohol-abusing characters by comedies and dramas. Interestingly, these top two show genres again differed significantly on theme appearance within the programs (see Table 5). Using one-way ANOVAs, comedies were significantly more likely not to show much impact on the nonalcohol-abusing characters, $F(1, 104) = 16.08, p < .05$, were significantly more likely to portray the drinking as humorous to the nonalcohol-abusing characters, $F(1, 104) = 12.42, p < .05$, were significantly less likely to portray the nonalcohol-abusing characters as embarrassed by the drinking, $F(1, 104) = 8.04, p < .05$, were significantly less likely to show the nonalcohol-abusing characters upset or distressed, $F(1, 104) = 30.07, p < .05$, and were significantly less likely to show the nonalcohol-abusing characters as being forced to take care of others, including the person drinking, $F(1, 104) = 21.32, p < .05$.

Finally, it is important to note from an E-E standpoint that Tables 4 and 5 show appearances of prohealth themes in the programs, as well as prodrinking themes. In Table 4, 8 of the top 11 reported themes

were prodrinking in nature. However, as shown in Table 5, participants reported more prohealth themes than prodrinking themes, with 55% of participants reporting that nonalcohol-abusing characters were upset or distressed by the drinking of others. Only 26% of participants indicated that the nonalcohol-abusing characters were not impacted much.

Discussion

In sum, as depicted by Tables 4 and 5 as well as the results showing significant differences between the themes portrayed in comedies versus dramas, this study indicates that college students receive both prohealth and prodrinking messages and images through the entertainment television they consume. Unfortunately, results shown in Table 4 suggest that prodrinking messages are still the most prevalent source of information regarding alcohol use portrayed in entertainment television. However, Tables 4 and 5 highlight the appearance of alcohol-related, prohealth themes as well. These findings are consistent with the authors' final hypothesis that although the majority of themes selected by participants would be prodrinking in nature, some E-E themes would be found within the programs.

Many programs, although more often dramas than comedies, provided a variety of prosocial and prohealth messages related to alcohol use (i.e., E-E). Some of the E-E themes identified within these programs addressed the effects of alcohol abuse on the user, the effects on nonusers, the consequences, as well as cues to action and general information regarding how to deal with alcohol abuse. Again, Bandura's social cognitive theory indicates that incorporation of these types of themes is essential when attempting to induce learning and behavior change in viewers (Bandura, 1986; Kennedy et al., 2004; Trull, 2005).

More specifically, the results concerning the themes pertaining to the impact on nonalcohol-abusing characters are evidence of E-E incorporation in entertainment television watched by college students. Although approximately 26% of the participants indicated that there was not much impact on the nonalcohol-abusing characters in the show and 16% indicated that the nonalcohol-abusing characters found the drinking humorous, the remaining "negative themes of impact" were identified as being present in a program by at least 14% of participants or more. The concept of the nonalcohol-abusing characters being impacted is useful for E-E because an awareness of how inappropriate behavior can impact others is often a good motivator for attitude or behavior change related to substance abuse (Fields, 2007; Rosenstock, 1974).

In addition to themes, results showed that other potentially effective E-E concepts were in use within

TABLE 5

Percentage of Participants Reporting Themes of Impact on Nonalcohol-abusing Characters

| Theme | Percentage Reporting Theme |
|-----------------------------|----------------------------|
| Not impacted much | 26 ^a |
| Upset/distressed | 55 ^b |
| Forced to care for others | 49 ^b |
| Annoyed by drinking | 34 |
| Embarrassed by the drinking | 32 ^b |
| Emotionally/mentally abused | 25 |
| Ignored the drinking | 18 |
| Were abandoned | 17 |
| Found the drinking humorous | 16 ^a |
| Were physically abused | 14 |

^aComedies significantly more likely to portray than dramas at $p < .05$

^bDramas significantly more likely to portray than comedies at $p < .05$

the listed programs. For example, 50% of the television programs that participants listed dedicated more than three episodes to the topic of alcohol abuse and 78% of participants reported watching the shows at least occasionally. Again, it is important in E-E to draw out a storyline over several episodes and to create appealing characters that keep the viewer interested (Brinson & Brown, 1997; Kennedy et al., 2004; Klimmt et al., 2006; Papa et al., 2000; Parrott, 1995).

Unfortunately, however, the findings from the present exploratory study also suggest that how alcohol use is portrayed within entertainment television programming should continue to be addressed. For example, despite the finding that participants perceived 68% of the characters to be alcoholics, only 32% of the characters were treated, and only 13% of treatments were perceived by participants as effective. Also, within comedies—one of the two most-frequently listed genres on the survey—only 17% of the characters were treated. The significant differences between comedies and dramas in this study suggest that comedies may be used less often as a source for E-E. Future research should continue to examine the realistic portrayal of alcoholism and treatment within all genres. If further research reports similar findings, producers of television programs should strongly consider the principles outlined within Bandura's social cognitive theory (Bandura, 1986, 2002) and Rosenstock's (1974) health belief model. These theories support the importance of airing more accurate depictions of alcoholism and the severity of the disease, susceptibility, the impacts on others, benefits of change, cues for action, and successful and realistic treatment.

Results from this study also showed that the more participants drank, the less of an issue they found alcohol abuse to be in society. This information is alarming in that the health-belief model indicates that individuals are only motivated to change their health-related behaviors if they perceive the effects as serious and themselves as vulnerable (Rosenstock, 1974). Applying the health-belief model to the results suggests that these participants will be less likely to engage in attitude or behavior change unless targeted specifically.

Fortunately, however, the analysis from the present study does indicate that carefully planned E-E campaigns have the potential to impact the heaviest-drinking group. For example, as partially predicted, individuals who drank 7 or more drinks in one sitting were significantly more entertained by the identified programs than the individuals in the other three groups. Also, as predicted, the episodes were significantly more likely to keep the attention of all three groups of drinkers when compared to the nondrinkers. In other words, the show was significantly more likely to keep the attention of the participants who drink alcohol than those

who do not drink. These data suggest that the heaviest drinkers are perhaps developing parasocial relationships (PSRs) with some of the characters on screen. Obviously the fact that the heaviest drinkers are more entertained and all drinkers are more attentive to the episodes discussing alcohol use indicates that, at the very least, some form of identification with the characters is occurring. If this is accurate, producers and researchers should continue to explore and monitor how best to present characters on television with whom college-aged drinkers can most easily relate, and who are most likely to influence them to drink responsibly.

In addition to monitoring the heaviest, and probably the most at-risk, drinking group, future research should continue to examine the general trends of E-E television programming on alcohol use as well as how E-E messages can successfully be transmitted through other media formats. Studies on movies and alcohol use have suggested that alcohol use in movies is more prevalent than in any other media (Brown & Walsh-Childers, 2002). Therefore, it is important to look at how E-E messages on alcohol use could be effectively integrated into other forms of media such as in movies, music, and the internet.

Despite the useful information gained from this study, future research should address some of its limitations. First, the present study used participants' autobiographical memory for the second section of the survey. In order to eliminate concerns about the accuracy of participants' memories, future researchers could conduct an experiment that shows specific, serialized E-E scenes to participants and then measures the impact on the participants. Additionally, future researchers should consider using a less-exploratory method by administering previously-created surveys with established norms.

If research on and application of E-E messages continues, it could play a significant role in helping to discourage alcohol abuse by college students. The present study has provided information about the use of alcohol by college students as well as how they perceive alcohol use. Specifically, the study provides useful insight into how college students perceive information about alcohol use as it is depicted in entertainment television. Finally, the results of the study provide an indication of the importance of Entertainment-Education programming.

References

- American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (Revised 4th ed.). Washington, DC: Author.
- Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Englewood Cliffs, NJ: Prentice Hall.
- Bandura, A. (2002). *Social cognitive theory of mass communication*. In J. Bryant & D. Zillmann (Eds.), *Media effects: Advances in*

- theory and research 2nd ed.* (pp. 121-153). Mahwah, NJ: Lawrence Erlbaum Associates.
- Brinson, S. L., & Brown, M. H. (1997). The AIDS risk narrative in the 1994 CDC campaign. *Journal of Health Communication, 2*, 101-112.
- Brown, J. D., & Walsh-Childers, K. (2002). Effects of media on personal and public health. In J. Bryant & D. Zillmann (Eds.), *Media effects: Advances in theory and research 2nd ed.* (pp. 453-488). Mahwah, NJ: Lawrence Erlbaum Associates.
- Brown, W. J., & Singhal, A. (1990). Ethical dilemmas of prosocial television. *Communication Quarterly, 38*, 268-280.
- Fields, R. (2007). *Drugs in perspective: A personalized look at substance abuse and abuse*. New York: McGraw-Hill.
- Harris, R. J. (2004). *A cognitive psychology of mass communication*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Hingson, R. W., Heeren, T., Zakocs, R. C., Kopstein, A., & Wechsler, H. (2002). Magnitude of alcohol-related mortality and morbidity among U.S. college students ages 18-24. *Journal of Studies on Alcohol, 63*, 136-144.
- Kennedy, M. G., O'Leary, A., Beck, V., Pollard, K., & Simpson, P. (2004). Increases in calls to the CDC National STD and AIDS hotline following AIDS-related episodes in a soap opera. *Journal of Communication, 54*, 287-301.
- Klimmt, C., Hartmann, T., & Schramm, H. (2006). Parasocial interactions and relationships. In J. Bryant & P. Vorderer (Eds.), *Psychology of entertainment* (pp. 291-313). Mahwah, NJ: Lawrence Erlbaum Associates.
- Knight, J. R., Wechsler, H., Kuo, M., Seibring, M., Weitzman, E. R., & Schuckit, M. A. (2002). Alcohol abuse and dependence among U.S. college students. *Journal of Studies on Alcohol, 63*, 263-270.
- Larson, M. S. (1991). Health-related messages embedded in prime-time television entertainment. *Health Communication, 3*, 175-184.
- Lozano, E. (1992). The force of myth on popular narratives: The case of melodramatic serials. *Communication Theory, 2*, 207-220.
- National Institute of Alcohol Abuse and Alcoholism. NIAAA council approves definition of binge drinking. *NIAAA Newsletter 2004*; No. 3, p. 3.
- Papa, M. J., Singhal, A., Law, S., Pant, S., Sood, S., Rogers, E. M., & Shefner-Rogers, C. L. (2000). Entertainment-education and social change: An analysis of parasocial interaction, social learning, collective efficacy, and paradoxical communication. *Journal of Communication, 50*, 31-55.
- Parrott, R. L. (1995). Motivation to attend to health messages. In E. Maibach & R. L. Parrott (Eds.), *Designing health messages: Approaches from communication theory and public health practice* (pp. 7-23). Thousand Oaks, CA: Sage.
- Petty, R., Priester, J. R., & Brinol, P. (2002). Mass media attitude change: Implications of the elaboration likelihood model of persuasion. In J. Bryant & D. Zillmann (Eds.), *Media effects: Advances in theory and research* (2nd ed.). (pp. 155-198). Mahwah, NJ: Lawrence Erlbaum Associates.
- Rogers, E. M., & Singhal, A. (1990). The academic perspective. In C. Atkin & L. Wallack (Eds.), *Mass communication and public health* (pp. 176-181). Newbury Park, CA: Sage.
- Rogers, E. M., Vaughan, P. W., Swalehe, R. M. A., Rao, N., Svenkerud, P., & Sood, S. (1999). Effects of an entertainment-education radio soap opera on family planning behavior in Tanzania. *Studies in Family Planning, 30* (3), 193-211.
- Rosenstock, I. M. (1974). Historical origins of the health belief model. In M. H. Becker (Ed.), *The health belief model and personal health behavior* (pp. 1-8). New Jersey: Charles B. Slack, Inc.
- Rosenzweig, J. (1999). Can TV improve us? *The American Prospect, 45*, 58-63.
- Sherry, J. L. (2002). Media saturation and entertainment-education. *Communication Theory, 12*, 206-224.
- Singhal, A., & Rogers, E. M. (1999). *Entertainment Education: A communication strategy for social change*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Trull, T. J. (2005). *Clinical psychology* (7th ed.). Belmont CA: Thomson Wadsworth.
- Wechsler, H., Lee, J. E., Kuo, M., & Lee, H. (2000). College binge drinking in the 1990's: A continuing problem. Results of the Harvard School of Public Health 1999 college alcohol study. *Journal of American College Health, 48*, 199-210.

APPENDIX

SURVEY

The following survey assesses how college students are impacted by the portrayal of alcohol use and abuse in entertainment television programs. Please answer the following questions as accurately and honestly as possible. **Please do not put your name or any identifying information on this sheet in order to ensure that your responses are completely anonymous.**

Age _____ Race /Ethnicity _____ Gender _____

On average, how often do you drink alcohol? (Fill in the blank spaces with the number of times you drink during one of the time frames listed or indicate if you no longer drink alcohol. If you do not drink at all, place a zero in the blank spaces.)

_____ times/week _____ times/month _____ times/year _____ I use to drink but quit

If you do drink alcohol, how many alcoholic drinks do you consume in one sitting/at one time?

1 2-3 4-6 7-9 10 or more

Do you feel alcohol abuse is an issue in our society?

Not at all 1 2 3 4 5 6 7 Very severe issue

How do you define alcohol abuse?

Do you know anyone personally who has/does abuse(d) alcohol? Yes No Not Sure

If yes, how could you tell they abuse(d) alcohol?

Think of an entertainment television show (comedy, drama, reality television, soap opera, or other) that portrayed alcohol abuse as a health-related issue for one or more of the characters. By "health-related" we mean that the alcohol use is somehow related to a character's physical, emotional, or mental state or to their general well being.

Name of television show: _____

Type of television show (Circle One):

Comedy/Sitcom Drama Reality TV Soap Opera Other _____

Do you watch this show regularly?

No Usually Not Occasionally More often than not Almost always

About how many episodes were dedicated to this alcohol-related topic (Circle one)?

1 2-3 4-5 6-7 8-9 10 or more Entire season

Overall, how entertaining did you find these episode(s)?

Not at all entertaining 1 2 3 4 5 6 7 Very entertaining

Overall, how well did these episode(s) keep your attention?

Not very well 1 2 3 4 5 6 7 Very well

About how long ago did you watch this/these episode(s)?

_____ Within last month _____ A few months ago _____ Last year _____ More than one year ago _____ Not sure

Gender of the character abusing alcohol?

Male Female Approximate age _____

How could you tell they abused alcohol?

Would you say the person abusing alcohol was an alcoholic?

Definitely yes Probably yes Not sure Probably not Definitely not

Was this character treated for alcohol abuse? Yes No Not Sure

If yes, how were they treated?

Was the treatment effective? Yes No Not Sure

How could you tell?

If the person was not treated, do you think they should have received treatment (Circle number)?

No 1 2 3 4 5 6 7 Yes

Why or why not?

What themes relating to alcohol were included in the program? (Check all that apply)

- | | |
|--|---|
| <input type="checkbox"/> Increase in awareness of alcohol abuse | <input type="checkbox"/> Drinking as a normal activity |
| <input type="checkbox"/> Treatment options for alcohol abuse | <input type="checkbox"/> Drinking as a fun recreational activity |
| <input type="checkbox"/> Emotional support for abusers | <input type="checkbox"/> Drinking to relieve stress |
| <input type="checkbox"/> Emotional support for relatives/friends of abuser | <input type="checkbox"/> Drinking to avoid responsibility |
| <input type="checkbox"/> Increase in knowledge about alcohol abuse | <input type="checkbox"/> Drinking to loosen up |
| <input type="checkbox"/> Change in attitude about alcohol abuse | <input type="checkbox"/> Drinking to forget |
| <input type="checkbox"/> Change in perception of who drinks | <input type="checkbox"/> Drinking to fit in socially |
| <input type="checkbox"/> Change in perception of why people drink | <input type="checkbox"/> Drinking to avoid emotions |
| <input type="checkbox"/> Suggestions for communicating about drinking | <input type="checkbox"/> Unplanned sexual activity |
| <input type="checkbox"/> Need for moderation when drinking | <input type="checkbox"/> Physical health issues as a result of drinking |
| <input type="checkbox"/> Need for designated drivers | <input type="checkbox"/> Emotional issues as a result of drinking |
| <input type="checkbox"/> Others? _____ | |

How were the nonalcohol-abusing characters of the program impacted? (Check all that apply)

- | | |
|--|---|
| <input type="checkbox"/> They were not impacted much | <input type="checkbox"/> They were annoyed by the drinking |
| <input type="checkbox"/> They were abandoned | <input type="checkbox"/> They were upset or distressed |
| <input type="checkbox"/> They found the drinking humorous | <input type="checkbox"/> They were emotionally/mentally abused |
| <input type="checkbox"/> They ignored the drinking | <input type="checkbox"/> They were physically abused |
| <input type="checkbox"/> They were embarrassed by the drinking | <input type="checkbox"/> They were forced to take care of others (including the person drinking) |
| <input type="checkbox"/> Other? _____ | |

Did the program provide any tips on getting additional information or support following the program? (For example, a hotline number or website) If so, what was provided

Was this program personally meaningful to you? If so, how?

What (if any) action(s) did you take following the program? (Check all that apply)

- | | |
|---|--|
| <input type="checkbox"/> Called hotline/visited website | <input type="checkbox"/> Encouraged others to seek help |
| <input type="checkbox"/> Intended to drink less | <input type="checkbox"/> Changed attitude/tolerance of alcohol abuse |
| <input type="checkbox"/> Actually drank less | <input type="checkbox"/> Thought about television show |
| <input type="checkbox"/> Used a designated driver | <input type="checkbox"/> No action |
| <input type="checkbox"/> Sought help | <input type="checkbox"/> Took another drink |
| <input type="checkbox"/> Other? _____ | |

What suggestions do you have as to how this program could be changed to better communicate the issue of alcohol abuse and to promote a reduction in alcohol abuse within society?

Sex Differences in 2D:4D Ratios, Aggression, and Spatial/Verbal Abilities

PATRICK COOPER

ROBERT MATCHOCK*

The Pennsylvania State University-Altoona Campus

The 2D:4D (second to fourth digit) ratio is thought to be a measure of prenatal testosterone exposure and is associated with traits such as aggression, sexual tendencies, and verbal/spatial intelligence. Fluctuating Asymmetry (FA) refers to deviation from bilateral symmetry caused by environmental problems during ontogeny. This study examined the relation among these factors. College students (N = 143) completed spatial and verbal fluency tests as well as aggression and sexual history questionnaires. Morphological measures of ears, ankles, and digit lengths were taken. The results indicate that 2D:4D ratios were sexually dimorphic with men having significantly lower ratios and higher spatial and aggression scores. Pearson correlations indicated that 2D:4D ratios were negatively correlated with physical aggression. The left hand ratios (L2D:4D) were negatively associated with aggression in women. Physical aggression and Total aggression were positively associated with spatial ability for men, but negatively associated with spatial ability for women. There were no sex differences for FA scores, verbal scores, and number of sexual partners. The 2D:4D ratios are discussed in terms of prenatal testosterone exposure.

Finger length ratios such as the ratio of the length of the second digit, or index finger (2D) divided by the length of the fourth digit, or ring finger (4D) serve as a developmental marker for prenatal testosterone (T) exposure. The 2D:4D ratios correlate with many behaviors (e.g., aggression, spatial ability, and number of sexual partners) and physical characteristics (e.g., fluctuating asymmetry) (Bailey & Hurd, 2005; Benderlioglu & Nelson, 2004; Csathó, et al. 2001). Past research suggests that lower ratios, closer to 0.9, are associated with higher levels of prenatal T and higher ratios, closer to 1.0, with lower levels of prenatal T. Prenatal T exposure and the 2D:4D ratio are consistently found to be a sexually dimorphic trait with men having lower ratios and higher levels of T (Manning, 2002; Putz, Gaulin, Sporter, & McBurney, 2004; Rahman, Korhonen, & Aslam, 2005).

Prenatal Testosterone Exposure and 2D:4D Ratios

Research suggests that prenatal androgen exposure, including T, is associated with variations in the 2D:4D ratio (Lalumière, Harris, & Rice, 2001). Examining children with congenital adrenal hyperplasia (CAH) supports this notion (Brown, Hines, Fane, & Breedlove, 2001). Children with CAH are exposed to excess andro-

gens during ontogeny. As a result, both sexes with CAH generally develop lower ratios compared to controls. This finding suggests that T could be responsible for the development of low 2D:4D ratios in both sexes in cases of CAH and provides support for the negative association between T and 2D:4D ratios in general.

Prenatal T is also related to *Hoxa* and *Hoxd* gene expression (Manning, 2002). These genes trigger the development of the digits and testes and suggest that, because T influences gonad development, it may also influence digit development. Manning, Bundred, Newton, and Flanagan (2003) suggests that the 2D:4D ratio is also affected by increased DNA replications of Cytosine-Adenine-Guanine (CAG) in the androgen receptor gene. Specifically, individuals with more CAG repeats have more T.

Correlations and Sex Differences in 2D:4D Ratios

Aggression is a sexually dimorphic trait with men consistently demonstrating higher levels of aggression than women (Bailey & Hurd, 2005). This sexual dichotomy includes several subtypes of aggression including physical aggression, verbal aggression, and hostility.

* Faculty mentor

Coupled with the fact that men also have significantly lower 2D:4D ratios, prenatal T exposure may explain the existence of higher aggression in men.

Spatial ability is also a sexually dimorphic trait that is associated with 2D:4D ratios (Csathó, et al., 2001; Kempel, et al., 2005; Sanders, Bereczkei, Csathó, & Manning, 2005). Similar to the organizational effects of T on digit formation and aggression, research suggests that prenatal T, as indicated by the 2D:4D ratio, might be related to spatial ability. Specifically, lower 2D:4D ratios would correlate with higher levels of spatial ability. Concerning observed sex differences in spatial ability, there is rarely a correlation between female 2D:4D ratios and the Mental Rotations Test (MRT; Vandenberg & Kuse, 1978), whereas there is a negative correlation for men.

2D:4D ratios may also correlate with verbal ability. On the Differential Aptitude Test (DAT), participants with higher right-hand 2D:4D ratios score higher on verbal analogies, which is consistent with the idea that *less* prenatal T (higher 2D:4D ratios) correlates with higher verbal ability. These results are supported by the fact that women with lower prenatal T and higher 2D:4D ratio score higher on verbal skills tests (Luxen & Buunk, 2005).

Evidence suggests that high levels of prenatal T adversely affect bilateral symmetry. Fluctuating asymmetry (FA), deviation from perfect bilateral symmetry, is associated with excess prenatal T exposure (Benderlioglu, Sciulli, & Nelson, 2004). Research shows that people find faces that are more symmetrical more attractive. From an evolutionary perspective, high FA individuals are considered less attractive and will be less likely to find a mate because symmetry is an indicator of health.

Evidence suggests that 2D:4D ratios also correlate with number of sexual partners and age of first sexual intercourse. Bogaert and Fisher (1995) found that number of sexual partners and early first sexual intercourse positively correlates with aggression, spatial ability, and activational (circulating) T. It is possible that 2D:4D ratios may also positively correlate with these sexual history traits.

The present study further examined the associations between 2D:4D ratios and aggression, spatial ability, verbal ability, FA, sexual history traits, as well as the inter-correlations among these variables. I predicted that individuals with higher prenatal T exposure, as indicated by 2D:4D ratios, would score higher on tests of aggression and spatial ability and score lower on tests of verbal ability, have increased number of sexual partners, and earlier age of first consensual sexual intercourse. I was also interested in the associations that FA has with these variables, particularly if high FA

is associated with increased levels of aggression and lower spatial ability, two traits associated with high prenatal T exposure.

Method

Participants

A total of 143 college students (61 men and 82 women) participated in the current study. The mean age and standard deviation of the sample was 22.9 years ($SD = 5.12$). For men, mean age was 22.11 years ($SD = 2.14$) and for women mean age was 23.51 years ($SD = 2.14$). Ethnic breakdown was: Caucasian (83.9%), African American (7.7%), Hispanic (3.5%), Asian (2.1%), and Native American (0.7%). Three participants (2.1%) marked their ethnicity as "other." Participants received course credit for various psychology courses. The local Institutional Review Board reviewed and approved all procedures.

Materials

A 20-item version of the Buss and Perry Aggression Questionnaire (1992) assessed participants' self-reported aggression. Sub-types of aggression include physical aggression (PA), Verbal Aggression (VA), Hostility (H), and Anger (A). A test of reliability for the questionnaire revealed that Cronbach's alpha was .89. Responses for each question ranged from 1 (*extremely uncharacteristic of me*) to 7 (*extremely characteristic of me*). Totaling all scores per aggression sub-type yielded a sub-type score (i.e., sum of all PA scores yields PA); totaling all aggression scores yielded a total aggression (TA) score.

The Thurstone Written Verbal Fluency Test (Thurstone & Thurstone, 1962) measured verbal fluency. The Vandenberg and Kuse Mental Rotations Test (MRT) measured spatial ability. A demographic questionnaire measured participants' age of first consensual sexual intercourse as well as other variables used in a larger study. To measure 2D:4D ratios, the experimenter photocopied participants' hands using a Xerox Workcentre Pro 35 photocopier. The experimenter measured each participant's ankle circumferences and ear lengths from top to bottom for a measure of FA. A composite or total FA score was calculated by summing the difference in ear lengths, ankle circumferences, and each corresponding digit length.

Procedure

Participants began the experiment by completing the demographic and aggression questionnaires. Neither questionnaire was timed; each participant could take as much time as needed to finish. Next, participants took the MRT ($M = 18.62$, $SD = 8.59$) and Verbal Fluency Test ($M = 82.11$, $SD = 20.47$).

I measured participants' ear lengths in millimeters (mm) using digital calipers, and each ankle circumference in millimeters (mm) using a standard tape measure exactly six inches above the sole of the foot. I measured second and fourth digits using the photocopy of the participant's basal crease to the tip of the finger. All photocopies were inspected for clear representations of the basal crease and the fingertip and measured by two independent measurers.

Results

Sex Differences in 2D:4D Ratios, FA, and Behaviors

The right hand ratio (R2D:4D) was significantly lower in men ($M = .959$, $SD = .032$) than in women ($M = .974$, $SD = .040$), $F(1, 139) = 5.30$, $p < .05$, but the left hand ratio (L2D:4D) was not (men: $M = .975$, $SD = .032$; women: $M = .986$, $SD = .045$). Left and right hand ratios were highly correlated for men, $r(59) = .403$, $p < .01$, and women, $r(78) = .551$, $p < .01$.

Total aggression was significantly higher in men ($M = 54.32$, $SD = 11.45$) than in women ($M = 43.65$, $SD = 13.54$), $F(1, 141) = 24.92$, $p < .001$. Of the subgroups of aggression, PA was significantly higher in men ($M = 13.64$, $SD = 4.19$) than in women ($M = 8.87$, $SD = 4.33$), $F(1, 141) = 42.572$, $p < .001$. In addition, H was also significantly higher in men ($M = 12.93$, $SD = 4.57$) than in women ($M = 11.14$, $SD = 4.15$), $F(1, 141) = 5.947$, $p < .05$. There were no significant sex differences in A.

Men ($M = 22.69$, $SD = 7.94$) scored significantly higher than women ($M = 15.50$, $SD = 7.76$) on the Mental Rotations Test, $F(1, 141) = 29.50$, $p < .001$. However, there were no sex differences in verbal ability, fluctuating asymmetry, or age of first consensual sexual intercourse. Using an ANCOVA with age as a covariate, the mean number of lifetime sexual partners was not significantly different between men ($M = 4.30$) and women ($M = 3.93$).

Correlations with 2D:4D Ratios, FA, and Behaviors

The R2D:4D ratio was negatively correlated with physical aggression for women, $r(79) = -.220$, $p < .05$, but not for men. The R2D:4D ratio did not correlate with any other aggression construct for either sex. For women, the L2D:4D ratio was negatively correlated with total aggression, $r(78) = -.273$, $p < .05$, negatively correlated with PA, $r(78) = -.239$, $p < .05$ and negatively correlated with A, $r(78) = -.254$, $p < .05$. Similar to the R2D:4D ratio findings, no aggression construct correlated with L2D:4D ratio for men. Neither the L2D:4D ratio nor the R2D:4D ratio correlated with spatial ability, verbal ability, fluctuating asymmetry, age of first sexual intercourse, or number of sex partners.

Fluctuating asymmetry positively correlated with

verbal aggression in women, $r(74) = .234$, $p < .05$, but not in men. FA did not correlate with any other aggression construct for either sex. It did, however, negatively correlate with spatial ability for men, $r(55) = -.266$, $p < .05$, but not women, $r(74) = -.152$, $p = .191$.

Spatial ability significantly correlated with total aggression for both sexes, albeit in different directions. For example, the correlation for men was positive, $r(60) = .283$, $p < .05$, but for women it was negative, $r(74) = -.240$, $p < .05$. Similarly, for the sub-type PA, its correlation with spatial ability for men was also positive, $r(60) = .339$, $p < .01$, but negative for women, $r(79) = -.273$, $p < .05$. Finally, spatial ability negatively correlated with A in women, $r(79) = -.277$, $p < .05$.

Verbal ability negatively correlated with anger, $r(78) = -.242$, $p < .05$ and hostility, $r(78) = -.229$, $p < .05$, in women. Verbal ability did not significantly correlate with any measured variables for men. A summary of

TABLE 1

Summary of Relevant Correlations Among Physical Aggression, Verbal Aggression, Total Aggression, Number of Sexual Partners, Spatial Ability, and Fluctuating Symmetry

| Correlations | Pearson's r | p-value |
|--|-------------|---------|
| 2D:4D Ratio Physical Aggression | $r = -.25$ | .003** |
| Total Aggression Number of Sexual Partners | $r = .20$ | .019* |
| Total Aggression (women) Spatial Ability | $r = -.24$ | .031* |
| Total Aggression (men) Spatial Ability | $r = .28$ | .026* |
| Physical Aggression (women) Spatial Ability | $r = -.27$ | .014* |
| Physical Aggression (men) Spatial Ability | $r = .34$ | .007** |
| Fluctuating Asymmetry (women) Verbal Aggression | $r = .23$ | .042* |
| Fluctuating Asymmetry (men) Spatial Ability | $r = -.27$ | .046* |

Note. * = statistically significant, $p < .05$; ** = $p < .01$

relevant correlations is shown in Table 1.

Discussion

Sex Differences in 2D:4D Ratios and Aggression

There were some expected and unexpected sex differences in the variables studied. As expected, 2D:4D ratio was sexually dimorphic, with men having lower ratios. Aggression and spatial ability were also sexually dimorphic; men had higher scores in each construct. However there were no sex differences for verbal ability, composite FA, age of first consensual sexual intercourse, or number of sex partners. Therefore, from these findings one cannot confidently assert that verbal ability, FA, age of first sexual intercourse, or number of sex partners are gender-differentiated.

In women, R2D:4D ratio and L2D:4D ratio negatively correlated with several subtypes of aggression. Although the literature is mixed on how digit ratios correlate with aggression in women (see Bailey & Hurd, 2005; Benderlioglu & Nelson, 2004), it may be that higher than normal levels of prenatal T exposure have a more robust or different effect on women because men normally experience high levels of prenatal T. Further research is needed to explore this hypothesis.

Aggression and Spatial Ability

Men scored significantly higher on aggression and spatial ability than women, consistent with past literature (Benderlioglu & Nelson, 2004; Fink, Manning, Williams, & Podmore-Nappin, 2007), but there was a differential effect for men and women. For men, high aggression scores were significantly associated with high spatial scores (positive correlation). For women, high aggression scores were significantly associated with low spatial scores (inverse correlation). Essentially, women who displayed high aggression did not perform well on spatial tasks. These quite different correlations are also suggestive of a U-shaped relation between aggression/testosterone and spatial ability. Future research should examine other possible causal pathways besides prenatal T exposure that mediate women's spatial ability.

Conclusions

This study examined the relation between 2D:4D ratios and behavioral characteristics. Results supported previous research that the 2D:4D ratio is indeed a sexually dimorphic trait. Furthermore, prenatal T exposure might have a profound influence on women, evidenced by the associations seen between 2D:4D ratios and aggression. However, one cannot firmly conclude this

pattern from the current study. Future experimental research should aim to delineate the effects of prenatal T exposure in women. The results also reveal a sex difference in the association between aggression and spatial ability, suggesting alternate pathways contribute to development of spatial ability in men and women.

References

- Bailey, A. A., & Hurd, P. L. (2005). Finger length ratio (2D:4D) correlates with physical aggression in men but not in women. *Biological Psychology*, *68*, 215-222.
- Benderlioglu, Z., & Nelson, R. J. (2004). Digit length ratios predict reactive aggression in women, but not in men. *Hormones and Behavior*, *46*, 558-564.
- Benderlioglu, Z., Sciuili, P. W., & Nelson, R. J. (2004). Fluctuating asymmetry predicts human reactive aggression. *American Journal of Human Biology*, *16*, 458-469.
- Bogaert, A. F., & Fisher, W. A. (1995). Predictors of university men's number of sexual partners. *The Journal of Sex Research*, *32*, 119-130.
- Brown, W., Hines, M., Fane, B., & Breedlove, S. (2001). Masculinized finger length ratios in humans with congenital adrenal hyperplasia (CAH). *Hormones and Behavior*, *39*, 325-326.
- Buss, A., & Perry, M. (1992). The aggression questionnaire. *Journal of Personality and Social Psychology*, *63*, 452-459.
- Csathó, Á., Osváth, A., Karádi, K., Bicsák, É., Manning, J., & Kállai, J. (2001). Spatial navigation related to the ratio of second to fourth digit length in women. *Learning and Individual Differences*, *13*, 239-249.
- Fink, B., Manning, J. T., Williams, J. H. G., & Podmore-Nappin, C. (2007). The 2nd to 4th digit ratio and developmental psychopathology in school-aged children. *Personality and Individual Differences*, *42*, 369-379.
- Kempel, P., Gohlke, B., Klempau, J., Zinsberger, P., Reuter, M., & Hennig, J. (2005). Second-to-fourth digit length, testosterone and spatial ability. *Intelligence*, *33*, 215-230.
- Lalumière, M. L., Harris, G. T., & Rice, M. E. (2001). Psychopathy and developmental instability. *Evolution and Human Behavior*, *22*, 75-92.
- Luxen, M. F., & Buunk, B. P. (2005). Second-to-fourth digit ratio related to verbal and numerical intelligence and the Big Five. *Personality and Individual Differences*, *39*, 959-966.
- Manning, J. T. (2002). *Digit ratio: A pointer to fertility, behavior, and health*. Piscataway, NJ: Rutgers University Press.
- Manning, J. T., Bundred, P. E., Newton, D. J., & Flanagan, B. F. (2003). The second to fourth digit ratio and variation in the androgen receptor gene. *Evolution and Human Behavior*, *24*, 399-405.
- Putz, D. A., Gaulin, S. J. C., Sporter, R. J., & McBurney, D. H. (2004). Sex hormones and finger length: What does 2D:4D indicate? *Evolution and Human Behavior*, *25*, 182-199.
- Rahman, Q., Korhonen, M., & Aslam, A. (2005). Sexually dimorphic 2D:4D ratio, height, weight, and their relation to number of sexual partners. *Personality and Individual Differences*, *39*, 83-92.
- Sanders, G., Bereczkei, T., Csathó, Á., & Manning, J. (2005). The ratio of the 2nd to 4th finger length predicts spatial ability in men but not women. *Cortex*, *41*, 789-795.
- Thurstone, L. L., & Thurstone, T. G. (Eds.). (1962). *SRA primary mental abilities*. New York: Oxford University Press.
- Vandenberg, S. G., & Kuse, A. R. (1978). Mental rotations, a group test of three-dimensional spatial visualization. *Perceptual and Motor Skills*, *47*, 599-604.

Influence of Name Cues to Ethnicity on Judgments of Guilt and Sentencing

JESSICA L. DAVISON

DANIELLE I. SCHNEIDER

SHEILA BROWNLOW*

Catawba College

We examined whether judgments of guilt and perceptions of criminality were influenced by subtle cues to the ethnicity of suspects: their names. Students read about the robbery-related arrest of two suspects—either William Blake/Frank Rogers or William Blake/Franco Rodriguez—and judged criminal intent and guilt, and also assigned each a sentence. Results revealed that women participants imposed slightly harsher punishments than men, but only when the accused pair included at least one Hispanic-named person. Moreover, women more than men were likely to see the accused as being more criminal, and perceptions of criminality were lower for the person named William. There were no name-related differences in sentencing or judgments of guilt. These findings suggest that when cues to egalitarian behavior are obvious, such as when determining a sentence length for two persons who are accused of the same crime, people exhibit little overt bias, but aversive bias is likely when judgments are not guided by explicit behavioral expectations.

Stereotypes are a way we categorize others when we have little knowledge about them (Devine, 1989). A simple salient characteristic such as sex, hair color, or race may be used to determine what another person is like. Whether this process is implicit (automatic) or explicit (controlled) depends on the situation and the motivation of the person making a judgment (Devine, Amodio, Harmon-Jones, Vance, & Plant, 2002). On one hand, when people are exposed to a prime about race, they may make implicit responses that are out of conscious control (Banaji, Hardin, & Rothman, 1993; Dovidio, Gaertner & Kawakami, 2002), leading to behavior that shows bias. On the other hand, people are motivated to avoid negative reactions from others and to retain a positive view of themselves. Therefore, when a response is easy to control people who have high internal motivation to be viewed as egalitarian will react without prejudice, showing that when people are capable and willing they can respond to others in social situations without using stereotypes (Devine et al., 2002).

One motivation to avoid acting in a biased manner is that doing so is not socially acceptable. However, social norms do not erase the fact that many people are uncomfortable when they interact with people of other

racess and ethnicities (Devine, 1989). Such discomfort is most likely among people who see themselves as egalitarian and who adhere to ideas of social, political, and economic equality, but who are nonetheless unsure of how to behave in social situations that include people of other races and ethnicities (Dovidio & Gaertner, 2000). Lack of overt bias in such situations does not, however, signal a lack of prejudice. Gaertner and Dovidio (1986) propose that bias in a subtle, indirect form has replaced overt bias, particularly racism, among White people who do not see themselves as prejudiced and who do not wish to act in a discriminatory way.

Aversive racism is a form of bias that has developed from the more blatant forms of racism that are a function of racial stereotypes (Gaertner & Dovidio, 1986). Aversive racism describes the ways that people who claim to have egalitarian views about race portray

Author note. The authors are grateful to the editor and the reviewers for their suggestions and help with the manuscript. Additionally, we thank Lauren Graham for assistance with our research. Portions of this work were presented at the 54th annual meeting of the Southeastern Psychological Association, Charlotte, NC, in March, 2008.

Correspondence regarding this research may be addressed to Jessica Davison at jessica.davison21@gmail.com

*Faculty mentor

themselves as non-prejudiced when their views or actions will be judged by other members of society; however when there is no fear of a negative social reaction, they express their prejudice (Dovidio et al., 2002). For example, White people will demonstrate prejudice to Black people by failing to offer help in ambiguous circumstances but only if it is not clear that social standards dictate that help should be given (Gaertner & Bickman, 1971; Gaertner & Dovidio, 1986). Similarly, Dovidio and Gaertner (2000) report that when it is difficult to judge how qualified a Black person is for a job because the applicant is neither clearly qualified nor clearly unqualified, White persons showed discriminatory behavior. However, when the person's qualifications for a job are clear, White people did not show discrimination. In sum, people show aversive racism when there is no pressure for them to act in a socially-acceptable manner. In such situations, people expose their true biases.

A potentially damaging place for aversive racism to manifest itself is in our judicial system. Race is a salient defendant quality affecting juror behavior, because jurors often do not have unlimited time or information about a defendant when making a decision (Steffensmeier & Demuth, 2002). There is ample evidence—from the laboratory to the courtroom, and for both judge and jury trials—that defendant race or ethnicity is related to judgments of culpability and sentencing. Studies of bias in judicial settings have focused primarily on comparisons between White and Black persons, showing, for example, that while judges are more punitive toward men, they are most punitive toward Black men (Steffensmeier & Hebert, 1999). Further, White persons are more likely than Black persons to receive probation rather than a jail sentence (Spohn, Gruhl, & Welch, 1981-1982; Unnever, Frazier, & Henrette, 1980), and the death penalty is sought more often for trials that include Black defendants (Paternoster, 1984).

The growth of the Hispanic population in the United States has allowed researchers to examine the possibility of judicial bias toward members of that group. Bodenhausen (1988) reported that people typically see Black and Hispanic persons as more likely than White persons to be guilty of a crime, and that they also are given lengthier sentences, perhaps justified by the expressed and explicit fear that those persons may commit similar crimes in the future. Muhlhausen (2004) and Steffensmeier and Demuth (2000) report that in actual trials, Hispanic defendants are given longer sentences than White defendants. The disparity in sentencing is greater among judges of color (Muhlhausen, 2004).

The severity of the crime, prior record, and the age and education of Black and Hispanic defendants may

be other contributing factors to sentence severity. Are Hispanic and Black people engaging disproportionately in criminal behavior and that is the reason for the harsher sentences? Or is it the case that we interpret information about crimes differently depending on who is being charged? Dane (1992) has noted that it is nearly impossible to eliminate all forms of stereotypes and biases from the courtroom, and that the influence of stereotypes can be seen in all aspects of the trial process. From simple defendant information such as socioeconomic status and residence, jurors will process subsequent information about the defendant in an automatic manner (Dane, 1992). Even being charged with a crime evokes specific pre-set biases about the type of behavior the accused person may engage in or may have been involved with in the past (Gordon, 1993). That is, if jurors believe that certain types of people commit certain classes of crime, they are more likely to vote for conviction if the accused is charged with a stereotype-relevant crime (Gordon, 1993). Perhaps juries respond in this manner because they do not fully evaluate the information given to them once they have enough information to make a decision (Bodenhausen, 1988, 1990).

While bias may be based on obvious information, it may, as noted earlier, also be based on tacit cues. In the latter case, people may not even realize that they are being primed to think a certain way about defendants, but they may show aversive racism. One subtle social cue is a name (Macrae, Mitchell, & Pendry, 2002), which can signal a person's class, gender, and age (Koole & Pelham, 2003). Names may also serve as cues to a person's race or ethnicity, activating our beliefs about the characteristics of a person and what sort of crime that person may or may not be likely to commit. For example, Shaneberger, Williamson, and Brownlow (1996) examined guilt and sentencing judgments of women accused of a crime, varying the names of the accused printed on a police report. One name was judged to belong to a Black woman, the other to belong to a White woman. Participants did not show explicit indications of racial bias, as they provided equal sentences and advocated equal treatment under the law. However, the participants perceived that the Black woman was more likely to have committed a similar crime in the past and that her behavior may have been affected by social forces, which could then allow them to justify any possible future biased behavior toward her or toward people in her social group. By establishing real concerns about the possibility that criminal behavior was part of a pattern, people have a plausible and defensible rationale to be more punitive.

In this study we examined how name cues to ethnicity affected beliefs about guilt and the sentencing

of persons with Hispanic names. “Name dropping” or other uses of name as a representation of race or ethnicity has not been used to examine bias toward Hispanics; therefore our research represents a new line of inquiry with regard to ethnicity and is an extension of name research focusing on comparisons of perceptions of Black and White people. We predicted that people would not show overt bias to Hispanic men accused of a crime by assigning them longer prison sentences, but that they would show subtle forms of bias in measures tapping other beliefs about the accused, such as the likelihood of past and future criminal behavior.

Method

Overview and Design

We randomly assigned men and women participants to one of two groups; in both cases they read a police report of a local crime committed by two men. The only difference in the police reports given to participants in these groups was whether the suspects listed on the report both had White/non-Hispanic names or whether one had a White/non-Hispanic name and one a Hispanic name. Participants evaluated likelihood of guilt and assigned a sentence to both accused persons, answered recall questions, and gave their personal background information. These manipulations resulted in a 2 x 2 x 2 (Participant Sex x Name Pair Mix: William—Frank or William—Franco x Name: William and either Frank or Franco) mixed design, with repeated measures on the last factor.

Participants

A total of 102 college students (52 men, 50 women; M age = 20.23) participated in this study. Most ($n = 88$) were White, and the remainder were Black ($n = 9$) or other ($n = 5$). Participants received course credit in psychology or sociology courses for their assistance.

Stimulus Material

Police report. We acquired a police report from the local police department and created an account of an armed robbery based on an actual crime noted in the *Police Blotter* of the local newspaper. The only difference in the report was that for half of the participants the suspects were listed as William Blake and Frank Rogers, and for the other half the suspects were William Blake and Franco Rodriguez.¹ The information given about the crime included the type of crime (armed robbery), date and time (September 5 at 11:30 pm), where the crime occurred (Rushco Convenience store) and the type of weapon used (hand gun). Information about the suspects included height (6 ft for both), weight (200 pounds; 145 pounds), age (25 years old for both) and their clothing (white shirt and blue jeans; black shirt and

black pants). No specification of suspect race and/or ethnicity was given in the report, and the physical characteristics listed (e.g., height, weight and clothing) could belong to either a White person or a Hispanic person. Potential personal information about the suspects, such as address, phone number, and social security number were blacked out with a marker on the report in order to lend credibility to the manipulation.

Dependent Measures

Likelihood of guilt and sentencing. Participants indicated the degree to which they agreed that each accused person was guilty, using 7-point scales with endpoints labeled 1 (*strongly disagree*) and 7 (*strongly agree*). After studying the police report, participants judged the likelihood that each of the suspects committed the crime, the likelihood that they had committed similar previous crimes, and the likelihood each would commit similar crimes in the future, each taken on a scale from 1 (*very unlikely*) to 7 (*very likely*). Additionally, they revealed their beliefs about how harshly the suspect should be punished from 1 (*as lenient as law allows*) to 7 (*as harshly as the law allows*). Lastly, participants sentenced the suspect or recommended probation, using sentencing guidelines which stipulated a 14-to-40 year sentence option for this crime (taken from the *North Carolina Criminal Law and Motor Vehicle Handbook*, 1993).

Recall. In order to ensure participants read and remembered the materials, participants answered seven questions about the crime after they completed the scales about guilt and sentencing. The type of crime and names of suspects were the most important of these questions, and if those two were not answered correctly (as occurred in eight instances), the data were discarded. The other questions dealt with location, age of the suspects, where they were apprehended, and the weapon used. We calculated recall accuracy as the percentage of correct answers to the seven questions about the crime.

Personal Background. Participants also provided personal information (age, sex, race, college class) and indicated whether any member of their family was an attorney or member of the police force. We collected this information because participants with exposure to the criminal justice system through family members

¹ A total of 14 students (all women) read several first names (taken from the police report section of the local newspaper) and indicated whether they thought the name-holder was White, Black, or Hispanic. The first name, Franco, was judged by all raters as belonging to a Hispanic person; the name Frank was seen by most ($n = 13$) as belonging to a White person, as was the name William ($n = 12$). None of the 37 names in this pretest were seen as *always* belonging to a White person, but William and Frank were deemed the best representative name for a White/non-Hispanic person.

may have better knowledge about how sentencing should occur.

Procedure

Up to 10 students took part in the experiment at one time. We first gave participants a packet that included directions and a description of the experimental procedure; once they indicated they understood the procedure and signed the consent form this packet was collected. We then distributed a packet that included the police report, followed by the questions about guilt and sentencing. The report was available for the participants' use while making those judgments. We collected the first packet of questions and distributed the second packet with the recall and background questions, which participants had to complete without the aid of the police report. After the whole group was finished, the experimenter explained that the general purpose of the study was to determine how people assigned guilt and punishment. We provided a complete description of the hypotheses, experimental procedure, and findings via e-mail at the end of the semester.

Results

Recall Manipulation Checks

A total of 94 (92%) students recalled the four most important questions about the police report: the names of each accused person, the type of crime, and where it was committed. Moreover, men (82.0%) and women (83.9%) recalled report information at a similar rate. Therefore, all but eight participants recalled enough information to be able to answer the subsequent questions.

Overview

Measures of guilt, sentencing, harshness of the penalty, perception of how much influence social forces contributed to the accused's behavior, and a composite measure of criminality were analyzed separately in 2 x 2 x 2 (Participant Sex x Name Pair Mix: William—Frank or William—Franco x Name: William and either Frank or Franco) mixed ANOVAs, with repeated measures on the last factor. The means and standard deviations from the analyses are located in Table 1. The criminality measure was a mean of three items that were strongly related and which may have tapped similar constructs: the chance the accused had done something similar in the past, the possibility that the accused would engage in worse crimes if left unchecked, and the likelihood that the accused might perform a similar crime in the future. These measures were highly correlated within name (Frank or Franco and William), all r 's(100) = .40 to .65, all p 's < .001.

Influence of Participant Sex, Name Mix of Suspects, and Names on Adjudications

Judgments of guilt, the recommended sentence length, and the likelihood that social factors were behind the actions of the accused persons were not influenced by any manipulations, although both the Suspect Name x Sex and Name Mix x Sex interactions for sentence length were marginally significant, both F 's (1, 98) = 2.79, both p 's = .10.

As hypothesized, more subtle measures of bias were influenced by the manipulations. Participants' assignment of the severity of the penalty for the crime was affected by names. For penalty harshness there was a significant interaction of Participant Sex and Suspect Name, $F(1, 98) = 4.67$, $p < .01$, partial $\eta^2 = .06$, and Participant Sex and Name Mix, $F(1, 98) = 4.07$, $p < .05$, partial $\eta^2 = .04$. Means from these interactions are displayed in Tables 2 and 3, respectively. For the Sex x Name interaction, post-hoc Scheffé tests ($\alpha = .05$) demonstrated that women participants ($M = 5.36$) gave a harsher penalty to William than did the men ($M = 4.87$), but that harshness toward Frank or Franco did not differ by sex (M 's = 5.23 and 5.06 for women and men, respectively). The Scheffé test examining sex differences showed a marginal trend ($p < .07$), as women were somewhat more harsh toward the accused men in the mixed-name condition ($M = 5.64$) versus the non-mixed name condition ($M = 4.96$). Men did not show differences in harshness toward the accused as a function of name-mix (M 's = 5.06 for non-Hispanic White men and 4.87 for the pair that included a mixture of names). No other main effects or interactions were significant on this measure.

Participant sex and name also influenced judgments of criminality. Specifically, the analysis produced a main effect of name, demonstrating that William was seen as less criminal ($M = 5.31$) than Frank or Franco ($M = 5.45$), $F(1, 98) = 8.83$, $p < .005$, partial $\eta^2 = .08$. Women judged the general criminality of the accused persons to be higher than did men, $F(1, 98) = 4.59$, $p < .05$, partial $\eta^2 = .05$. The main effect for Name Mix approached significance, $F(1, 98) = 2.79$, $p = .10$, as did the Name x Name Mix interaction, $F(1, 98) = 3.04$, $p = .08$. No other main effects or interactions approached significance.

Ethnicity/Race and Family Background

In an effort to rule out alternative possibilities that these results were a function of participant race or participant family background, the criminality ANOVA was recalculated for only White participants ($n = 88$; 42 men, 46 women) and again for those participants who did not have family in law enforcement ($n = 89$; 45 men and 44 women). In both ANOVAs the main effect for name

remained significant (i.e., William less criminal than Frank or Franco), as it was in the ANOVAs including all the participants. The effect for participant sex (i.e., women more punitive) was marginally significant ($p = .06$). Therefore, removing participants who had family in the criminal justice system or who were non-White did not change the pattern of the criminality data.

Discussion

The results of this study demonstrated that the ethnicity of an accused man’s name did not influence explicit measures of guilt or perceptions of how lengthy his sentence should be. However, participants showed some evidence of aversive bias. In general, perceptions of criminality were lower for the person named William as opposed to his purported partner in the criminal act, who was named *either* Frank or Franco. There was some evidence of “guilt by association” for the accused person named William, in that women participants were more likely than men to advocate harsher punishments, particularly when the accused pair included at least one Hispanic-named person. However, women participants were more likely than men to see the accused persons as being more criminal in general.

Criminality judgments encompassed the likelihood that the suspect had engaged in criminal behavior in the past, would do so in the future, and could possibly commit worse crimes. In judgments of criminality, participants may have demonstrated more aversive bias because salient social cues were not present to dictate behavior: it would have been clear bias for people to assign Franco a longer sentence than William, but not to indicate greater concern about Franco’s future criminal potential. This pattern is similar to that reported by Shaneberger et al. (1996), whose study of Black and White women accused of a crime also produced no differences in sentencing between Black and White accused persons, but whose data revealed that participants believed the Black woman was more likely to have a criminal background, which could then allow people to justify any possible future biased behavior toward her or toward people in her social group.² Perhaps women in this study found both men (but especially those who were in mixed-ethnicity pairs)

² A similar pattern of data has been revealed in another study conducted among members of the same participant population, as women were more punitive in general toward accused persons (Graham, Capano, Davison, Schneider, & Brownlow, 2009).

TABLE 1

Means and Standard Deviations for Measures of Guilt and Culpability According to Participant Sex, Name Mix, and Name

| | Participant Sex | | | | | | | |
|--------------------------|-------------------|---------------------|--------------------|---------------------|-------------------|---------------------|--------------------|-----------------|
| | Men | | | | Women | | | |
| | Name Mix | | | | | | | |
| | All White | | Mixed | | All White | | Mixed | |
| William (n = 26) | Frank (n = 26) | William (n = 26) | Franco (n = 26) | William (n = 24) | Frank (n = 24) | William (n = 26) | Franco (n = 26) | |
| Guilt | 5.27 (1.34) | 5.19 (1.44) | 5.35 (1.09) | 5.42 (1.14) | 5.38 (1.01) | 5.42 (0.97) | 5.65 (0.94) | 5.30 (1.23) |
| Sentence | 9.94 (8.88) | 9.44 (8.69) | 9.00 (6.29) | 8.56 (6.32) | 7.31 (5.90) | 7.67 (5.97) | 11.23 (7.03) | 11.31 (7.14) |
| Harshness ¹ | 4.96 (1.31) | 5.15 (1.41) | 4.77 (1.11) | 4.96 (1.04) | 4.96 (0.81) | 4.96 (0.86) | 5.73 (1.08) | 5.54 (1.21) |
| Social | 5.23 (1.70) | 5.27 (1.77) | 5.38 (1.06) | 5.38 (1.20) | 5.63 (1.06) | 5.63 (0.97) | 5.69 (1.23) | 5.88 (1.18) |
| Criminality ² | 5.04 (1.30) | 5.21 (1.02) | 5.26 (0.86) | 5.38 (0.88) | 5.36 (0.75) | 5.45 (0.86) | 5.62 (0.83) | 5.91 (0.85) |

Note. ¹Harshness of penalty responses showed significant Sex x Name and Sex x Name Mix interactions, p 's < .05. ²Criminality ratings showed significant effects of name and participant sex, p 's < .05.

to be more criminal due to their implicit stereotypes of criminals, but did not sentence the two differently because they used controlled processes that took into account the possibility that other people would see them as prejudiced if they assigned a longer sentence to Franco (Devine & Elliot, 1995; Dovidio et al., 2002).

Participants may have employed automatic processing when asked about personal characteristics of the suspect due to the fact that making such judgments would not be overtly biased. Therefore, their implicit beliefs and their explicit beliefs may not have mirrored each other.

TABLE 2

Means and Standard Deviations from Participant Sex x Accused's Name Interactions for Judgments of Criminality and Harshness of Sentence.

| | Participant Sex | | | | | |
|-------------|-----------------------------|----------------|-----------------------------|-----------------------------|----------------|-----------------------------|
| | Men (<i>n</i> = 52) | | | Women (<i>n</i> = 50) | | |
| | Name of Accused | | | | | |
| | William | Frank/co | Total | William | Frank/co | Total |
| Harshness | 4.87 _a (0.15) | 5.06 (0.16) | 4.96 (0.11) | 5.36 _b (0.26) | 5.23 (0.16) | 5.30 (0.15) |
| Criminality | 5.15 (0.13) | 5.22 (0.14) | 5.18 _a (0.13) | 5.41 (0.19) | 5.76 (0.19) | 5.59 _b (0.14) |

Note. Means marked with different subscripts within rows differ significantly, $p < .05$. Additionally, there was a main effect for name of accused, with judgments for William ($M = 5.31$) lower than those for Frank/co ($M = 5.45$).

TABLE 3

Means and Standard Deviations from Participant Sex x Name Mix Interactions for Judgments of Criminality and Harshness of Sentence.

| | Participant Sex | | | | | |
|-------------|-------------------------|----------------|-----------------------------|-----------------------------|-----------------------------|-----------------------------|
| | Men (<i>n</i> = 52) | | | Women (<i>n</i> = 50) | | |
| | Name of Accused | | | | | |
| | All White | Mixed | Total | All White | Mixed | Total |
| Harshness | 5.06 (0.21) | 4.87 (0.21) | 4.96 (0.22) | 4.96 _a (0.21) | 5.64 _b (0.21) | 5.30 (0.15) |
| Criminality | 5.05 (0.13) | 5.32 (0.14) | 5.18 _a (0.13) | 5.41 (0.19) | 5.75 (0.18) | 5.59 _b (0.14) |

Note. Means with different subscripts within rows differ at $p < .05$, except those two means within women participants, $p < .07$.

There are three plausible explanations for women's tendency to be more greatly influenced by the name cues than the men participants. First, women may have evaluated the information in the police report differently, thinking it was acceptable to be harsh because they read through all the background information and thus became concerned about criminality, which provided a somewhat reasonable justification for their beliefs (Bodenhausen, 1988, 1990; Khan & Lambert, 2001). This explanation is supported by research that shows that women are more likely than men to worry that a criminal may be involved in future legal trouble (Velin & Walters, 1988). Further bias would then be possible, as the argument about criminality might afford women an external attribution for their beliefs, allowing them to continue to perceive themselves as egalitarian. A second explanation for the data is that men are simply less punitive than women, a finding consistent with other research (e.g., Lee, Lee, Horowitz, & King, 2006) showing women are harsher than men toward accused persons of color. Third, the presence of the gun in the crime may have been more alarming to women than to men. Generally, crimes involving weapons garner longer sentences, and the more salient that a gun was used as a weapon in the crime, the harsher the sentence, particularly for women (Dienstbier et al., 1998).

The results also show the viability of using names as a subtle cue to ethnicity, which may in turn allow for a more realistic appraisal of people's automatic and controlled processing. That is, people often read the newspaper and make judgments about the guilt or innocence of persons accused of crimes. However, we did not ask participants to judge the ethnicity/race of the persons in the police report after they had completed other recall questions, so we do not know whether participants would have been able—or willing—to guess ethnicity. Although collecting this information is essential for future studies, the likelihood that participants would have acknowledged that the information affected their decisions, even if it did, would be low (Pennington & Hastie, 1990).

Adjustments to our method in subsequent studies may shed light on several questions that arose from the data. As noted, asking participants to judge the ethnicity of the accused persons is necessary, as is determining why participants did not use sentencing guidelines in their responses. The sentencing guidelines indicated a 14-to-40 year window, but participants averaged a nine and one-half year sentence for the crime. An assessment of either egalitarianism or its opposite, social dominance, might help determine which person variables interact with cues to race and ethnicity to influence judgments of others. To improve the external validity of the study, a more

diverse sample of participants that includes people of varied ethnic, residential, and socioeconomic backgrounds is needed to determine whether the name cues affect people the same way. More importantly, a more realistic jury situation—where people make yes or no judgments of guilt while working with other persons—would also increase the generalizability of the results, as estimating the degree to which someone appears guilty is not an option in the judicial system. Despite needed improvements to the method, this study adds a dimension to understanding how members of one of the largest and growing minority groups in America may be viewed differently owing to a cue to their origin. Given that the U.S. Census Bureau (2009) reports that Hispanics currently comprise around 15% of the U.S. population (a figure expected to increase to about 20% by 2030) and as the ethnic and racial composition of America continues to shift and as the nation becomes more diverse, an understanding of the origins and outcomes of our attitudes as they influence our behavior in the judicial system becomes increasingly important.

References

- Banaji, M. R., Hardin, C., & Rothman, A. J. (1993). Implicit stereotyping in person judgment. *Journal of Personality and Social Psychology, 65*, 272-281.
- Bodenhausen, G. V. (1988). Stereotypic biases in social decision making and memory: Testing process models of stereotype use. *Journal of Personality and Social Psychology, 55*, 726-738.
- Bodenhausen, G. V. (1990). Second guessing the jury: Stereotypic and hindsight biases in perceptions of court cases. *Journal of Applied Social Psychology, 20*, 1112-1121.
- Dane, F. C. (1992). Applying social psychology in the courtroom: Understanding stereotypes in jury decision making. *Contemporary Social Psychology, 16*, 33-36.
- Devine, P. G. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology, 56*, 5-18.
- Devine, P. G., Amodio, D. M., Harmon-Jones, E., Vance, S. L., & Plant, E. A. (2002). The regulation of explicit and implicit race bias: The role of motivations to respond without prejudice. *Journal of Personality and Social Psychology, 82*, 835-845.
- Devine, P. G., & Elliot, A. J. (1995). Are racial stereotypes really fading? The Princeton Trilogy revisited. *Personality and Social Psychology Bulletin, 21*, 1139-1150.
- Dienstbier, R. A., Roesch, S. C., Mizumoto, A., Hemenover, S. H., Lott, R. C., & Carlo, G. (1998). Effects of weapons on guilt judgments and sentencing recommendations for criminals. *Basic and Applied Social Psychology, 20*, 93-102.
- Dovidio, J. F., & Gaertner, S. L. (2000). Aversive racism and selection decisions: 1989 and 1999. *Psychological Science, 11*, 315-319.
- Dovidio, J. F., Gaertner, S. L., & Kawakami, K. (2002). Implicit and explicit prejudice and interracial interaction. *Journal of Personality and Social Psychology, 82*, 62-68.
- Gaertner, S. L., & Bickman, L. (1971). Effects of race on the elicitation of helping behavior: The wrong number technique. *Journal of Personality and Social Psychology, 20*, 218-222.
- Gaertner, S. L., & Dovidio, J. F. (1986). The aversive form of racism. In A. G. Halberstadt & S. L. Ellyson (Eds.), *Social psychology readings*, (pp. 270-280). New York: McGraw-Hill.
- Gordon, R. A. (1993). The effect of strong versus weak evidence on the assessment of race stereotypic and race nonstereotypic crimes. *Journal of Applied Social Psychology, 23*, 734-749.

- Graham, L. E., Capano, K. M., Davison, J. L., Schneider, D. I., & Brownlow, S. (2009, February). *Do name cues to ethnicity affect judgments of criminality?* Poster presented at the 55th annual meeting of the Southeastern Psychological Association, New Orleans, LA.
- Khan, S. R., & Lambert, A. J. (2001). Perceptions of rational discrimination: When do people attempt to justify race-based prejudice? *Basic and Applied Social Psychology, 23*, 43-53.
- Koole, S. L., & Pelham, B. W. (2003). On the nature of implicit self-esteem: The case of the name letter effect. In S. J. Spencer & S. Fein (Eds.), *Motivated social perception: The Ontario symposium on personality and social psychology* (Vol. 9, pp. 93-116). Mahwah, NJ: Erlbaum.
- Lee, R. F., Lee, L. F., Horowitz, I. A., & King, E. (2006). The effects of defendant race, victim race, and juror gender on evidence processing in a murder trial. *Behavioral Sciences and the Law, 24*, 179-198.
- Macrae, C. N., Mitchell, J. P., & Pendry, L. F. (2002). What's in a forename? Cue familiarity and stereotypical thinking. *Journal of Experimental Social Psychology, 38*, 186-193.
- Muhlhausen, D. B. (2004). The determinants of sentencing in Pennsylvania: Do the characteristics of the judge matter? *Report of the Heritage Center for Data Analysis*. Washington, DC: Heritage Foundation.
- North Carolina Criminal Law and Motor Vehicle Handbook*. (1993). Melbourne, FL: Gould Publications.
- Paternoster, R. (1984). Prosecutorial discretion in requesting the death penalty: A case of victim-based racial discrimination. *Law and Society Review, 18*, 437-478.
- Pennington, N., & Hastie, R. (1990). Practical implications of psychological research on juror and jury decision making. *Personality and Social Psychology Bulletin, 16*, 90-105.
- Shaneberger, D., Williamson, N., & Brownlow, S. (1996). ...And justice for all? The effect of name cues to race on judicial decisions. *Psi Chi Journal of Undergraduate Research, 1*, 2-9.
- Spohn, C., Gruhl, J., & Welch, S. (1981-1982). The effect of race on sentencing: A re-examination of an unsettled question. *Law and Society Review, 16*.
- Steffensmeier, D., & Demuth, S. (2002). Ethnicity and sentencing outcomes in US federal courts: Who is punished more harshly? *American Sociological Review, 65*, 705-729.
- Steffensmeier, D., & Hebert, C. (1999). Women and men policymakers: Does the judge's gender affect the sentencing of criminal defendants? *Social Forces, 77*, 1163-1196.
- Unnever, J. D., Frazier, C. E., Henrette, J. C. (1980). Race differences in criminal sentencing. *The Sociological Quarterly, 21*, 197-205.
- United States Census Bureau (2009). *Hispanic Population in the United States*. Retrieved June 29, 2009, from http://www.census.gov/population/www/socdemo/hispanic/hispanic_pop_presentation.html
- Velin, R. A., & Walters, H. A. (1988). Intentional evidence: Effects on mock-jurors' perceptions. *Perceptual and Motor Skills, 62*, 290.

Could Attempts to Reduce the Effects of Inadmissible Evidence Go Too Far?

JUSTIN JOSEPH

John Carroll University

The present research explored whether jurors might convict defendants less frequently when a judge warns the jurors that inadmissible evidence could be presented at trial and later instructs them to disregard that evidence because it is unreliable. The study also attempted to determine if jurors in either the previous condition or a second condition in which a judge instructed the jurors to disregard the inadmissible evidence but did not give the later warning, would exhibit the backfire effect by convicting the defendant more often. The results indicated that the judge's admonishment and warning eliminated the backfire effect, but the conviction rates in the condition with a warning were not lower than the control condition in which the evidence was not presented.

Numerous studies have shown a significant relationship between inadmissible evidence and the decisions that juries make (see Steblay, Hosch, Culhane, and McWethy, 2006, for a review of the literature). Much of the research on the effects of inadmissible evidence involves a phenomenon called the backfire effect. Under current legal guidelines, if jurors hear inadmissible evidence in the courtroom, the judge may instruct the jurors to disregard that evidence. However, previous research suggests that jurors are unable to disregard this evidence during their decision-making process. To the contrary, the presence of inadmissible evidence causes jurors to convict more often than if the same evidence is admitted in the trial as admissible evidence (Brewer & Williams, 2005).

Inadmissible evidence is evidence that a judge determines is unreliable or irrelevant, or the potential of the evidence to prejudice the jurors outweighs its evidentiary value (Brewer & Williams, 2005). For example, a judge may not allow a suspect's prior convictions into a trial because the jurors would be greatly prejudiced by that information, which has little evidentiary value for the current case. In other words, a judge may rule a piece of testimony inadmissible because the jurors may not appropriately regard the evidentiary value of that evidence.

Pickel (1995), Sommers and Kassin (2001), and Wissler and Saks (1985) demonstrated the in-trial biasing effects of inadmissible evidence. Although the

judge may deem evidence inadmissible before the trial, eyewitnesses or attorneys may still present it in trial as a matter of strategy or by mistake. Pickel (1995) and Sommers and Kassin (2001) presented prior convictions as inadmissible evidence to mock jurors and then instructed them to disregard the evidence for determining their verdict. The results demonstrated that mock jurors cannot disregard the inadmissible evidence. In a similar study, Wissler and Saks (1985) presented prior convictions as inadmissible evidence. The mock jurors in their study also exhibited higher conviction rates when the judge instructed them to ignore the inadmissible evidence. In both studies, the mock jurors exhibited the backfire effect because the jurors who heard the inadmissible evidence followed by an admonishment demonstrated higher conviction rates than jurors who were presented with the evidence and not instructed to disregard it.

Lee, Kraus, and Lieberman (2005) focused their study on hearsay inadmissible evidence and also found the backfire effect. Hearsay evidence is deemed inadmissible in courts because it is a statement from a speaker who is not in the court. The validity of the hearsay statement cannot be determined during the trial. This study also confirmed the presence of the backfire effect in civil trials, whereas most of the previous research focused on the backfire effect in criminal

Faculty mentor: Janet D. Larsen

trials. Other studies also demonstrated the backfire effect with various types of evidence and various types of cases (Cook, Arndt, & Lieberman, 2004; Fein, McCloskey, & Tomlinson, 1997; Kassin & Sommers, 1997; Kassin & Sukel, 1997; Lee, Kraus, & Lieberman, 2005). Therefore, the backfire effect occurs with many different kinds of inadmissible evidence and there is little difference between the backfire effect in civil and criminal cases (Kassin & Sukel, 1997; London & Nunez, 2000; Pickel, 1995).

There are also numerous proposed explanations and solutions for the backfire effect. Reactance theory (Brehm, 1966) is one popular explanation of the backfire effect. Reactance theory explains the backfire effect as jurors feeling that their freedom is restricted when they receive instructions that certain evidence is inadmissible, and they should disregard it. Thus, jurors are more likely to use the inadmissible evidence in determining their verdict because their freedom is very important to them. Pickel (1995) successfully demonstrated reactance theory. In that study, the judge admonished jurors more harshly in one experimental condition than the other. The results yielded greater backfire effects in the condition in which the judge admonished the jurors more harshly. Though this theory is plausible and grounded in research, it paints a negative picture of human behavior. This theory simply states that the jurors want to counteract or rebel against the authority of the judge. The major weakness of reactance theory is that it does not take into account whether the jurors believe the evidence is reliable. It could be that these studies found greater backfire effects with stronger admonitions because the jurors simply believed the evidence was more important when the judge admonished them strongly and felt determined to do justice despite the justice system.

The pursuit of justice theory does take into account the jurors' feelings about the admissibility of the evidence. According to this theory, the backfire effect occurs because jurors believe that what they deem to be important information should be considered in their determination of guilt, regardless of the judge's admonitions. In other words, the jurors feel that it is their duty to decide which evidence should be admitted and that they are obligated by an inherent pursuit of justice to make their own determinations about the relevance of evidence (Fein, McCloskey, & Tomlinson, 1997; Kassin & Sommers, 1997; Sommers & Kassin, 2001). According to this theory, the backfire effect increases significantly if technical legal reasoning and not the substantive unreliability of the evidence is the reason for the exclusion of the evidence from the trial. The jurors demonstrate a certain level of skepticism about the fairness of the justice system and attempt to

reach the most just verdict on their own.

Another closely related theory explaining the backfire effect is that jurors discount the evidence only if the evidence seems suspicious to them. Otherwise, if they think the evidence is reliable, they disregard the admonition of the judge (Fein, McCloskey, & Tomlinson, 1997). This theory is known as the discounting and suspicion theory. In testing this theory, researchers made the jurors aware of the reasons for the inadmissibility of the evidence. One group was told that the information was unreliable, a substantive reason for denying admission of the evidence, and the other was told that it was obtained illegally, a technical legal reason. The mock jurors were more likely to disregard evidence when they believed it was inadmissible because it was unreliable. The jurors were also more likely to disregard evidence when they were made suspicious of the reliability of the evidence before it was presented.

Courts have tried at times to reduce the backfire effect, mostly through admonishment of the evidence after its presentation, but studies continue to show that these attempts are ineffective. Some attorneys have even been accused of using the backfire effect to their advantage because of the courts' ineffective methods of nullifying it. The legal system's method of nullifying the effects of inadmissible evidence changed little as a result of psychological research, as is demonstrated by the U.S. Supreme Court's decision in *Arizona v. Fulminante*. In this case, an inadmissible coerced confession was accidentally admitted into the trial evidence. The court ruled that this was a "harmless error" (Kassin & Sukel, 1997). Kassin and Sukel tested the Supreme Court's claim empirically and found that, even when the jurors know the confession is coerced, they still are unable to ignore it. Thus, the admission of inadmissible evidence into the *Fulminante* case was likely not a harmless error as the court decided. This is just one example of the court system's failure to make changes in procedure based on psychological research.

According to past research, explaining to jurors why the evidence is unreliable and giving a substantive rather than technical reason for disregarding the evidence can reduce the backfire effect experimentally. Warning the jurors before the trial that inadmissible evidence may be presented and that they should disregard it also reduces the backfire effect. Therefore, the hypothesis of the current research was that combining these methods from past research should create overcompensation in reducing the backfire effect and lead to lower than normal conviction rates of the defendant. This research combined the approaches used by Fein, McCloskey, and Tomlinson (1997) and Kassin and Sukel (1997). In the first condition, mock jurors read a trial transcript omitting any inadmissible evidence. In

the second condition, the mock jurors read that some wiretap evidence is inadmissible because it is unreliable. The third experimental condition remained largely the same as the second except, based on the findings of Schul (1993), jurors read a warning about the unreliability of inadmissible evidence before they read the trial transcript. In the fourth condition, participants read the transcript including the inadmissible evidence, with no indication that it was inadmissible. Thus, in the fourth condition the mock jurors read the inadmissible evidence as admissible evidence.

The first hypothesis of the current study was that warning jurors and giving a substantive reason for the inadmissibility of the evidence would not only correct for the backfire effect, but also would cause jurors to overcompensate and demonstrate lower conviction rates than in the control condition (which did not include the presentation of inadmissible evidence). That is, the mock jurors would exhibit the reverse of backfire effect. Secondly, I expected convictions in all cases to be lower than in the fourth condition when the mock jurors read the inadmissible wiretap evidence as admissible evidence. Another hypothesis was that the participants' level of confidence in their verdicts would be highest when the mock jurors read the inadmissible wiretap evidence as if it were admissible and lowest in the control condition with no wiretap evidence. The last hypothesis was that no difference would be found between the confidence levels of the two conditions in which the mock jurors read inadmissible evidence.

Method

Participants

Ninety-two students enrolled in courses at a small midwestern university participated in the study. Those students enrolled in introductory psychology courses received credit toward their introductory psychology class requirements for participation in the study. The researcher followed the APA code of ethics (American Psychological Association, 2002).

Materials

The participants read the experiment through *Media Lab* computer software (Emperisoft, New York). The present research used a trial transcript adapted from previous studies in the field (Fein, McCloskey, & Tomlinson, 1997; Kassin & Sukel 1997). The participants read the trial transcript entirely unaltered for the first condition (control condition). However, the trial transcript differed in a number of ways for the remaining conditions. For the other three conditions, jurors read inadmissible evidence added to the transcript, as used in previous research (Kassin & Sommers 1997). This involved the addition of a wiretap conversation

into the evidence presented (see Appendix). In the second condition (inadmissible no-warning condition), jurors read instructions to disregard the inadmissible wiretap evidence only after the presentation of the inadmissible evidence and they read no warning prior to the beginning of the transcript. The third condition (inadmissible warning condition) included a warning that inadmissible evidence could be in the trial and should be disregarded. This condition also included an admonition to disregard the inadmissible evidence following its presentation. In the fourth condition (admissible condition), the jurors read the transcript with the inadmissible wiretap evidence included, but it appeared to them as admissible evidence.

Procedure

A computer program determined the random assignment of participants to one of the four conditions: control, inadmissible no-warning, inadmissible warning, or admissible. The participants first read a screen instructing them that continuing in the experiment constituted their consent to participate. They then read instructions telling them to be as attentive as possible during the presentation of the trial transcript. In the warning condition, the jurors then read a warning at the beginning of the trial transcript:

THE COURT: I would like to forewarn the jury that there may be evidence presented here today that has been deemed inadmissible by the court, because it was unreliable. If this should occur I will instruct you to disregard the evidence and it should have no effect on your verdict.

In both the inadmissible no-warning and the inadmissible warning conditions the judge in the transcript admonished the jurors, following the presentation of the inadmissible evidence, to disregard it. Following an objection by the defense attorney to the inadmissible wiretap evidence the participants in the warning condition also saw a statement that read:

THE COURT: Sustained. I would ask the jury to disregard the tape conversation as the evidence has been deemed inadmissible. It is inadmissible because the tape was barely audible and it was too difficult to determine what was said. As I stated in the opening instructions inadmissible evidence is to be completely disregarded by the jury, because this evidence has been determined to be unreliable.

The participants in the inadmissible condition did not read the warning at the beginning of the trial and saw the same statement after the presentation of

the evidence without the last sentence referencing the warning at the beginning of the trial. Participants in the admissible condition read the wiretap evidence with no objection from the defense. Participants in the control condition did not read the wiretap evidence.

The computer program presented the trial transcript in sections ranging from 413-480 words. Each section remained on the screen for 1.5 min. After the allotted time, the participants could advance to the next screen by pressing a key when they finished reading that section of the transcript. After the participants read the transcript, they indicated whether they would find the defendant not guilty or guilty and rated their confidence in their verdict. After indicating their verdict, participants responded to several questions about the trial. They rated how well the prosecution and the defense presented their cases and the strength of the evidence for each side. They also rated the ethicality of the prosecution and the defense during the trial and the influence of various pieces of evidence on their verdict. The presentation of these questions was in a different random order for each participant and all responses were rated on a scale of 1 (*not at all confident*) to 7 (*very confident*).

Results

Verdict Measures

A chi-square test revealed that the type of transcript the participants read did not affect the probability of participants' verdicts being guilty or not guilty (see Table 1 for the percentage of guilty and not guilty verdicts in each condition).

A 2 (verdict: guilty or not guilty) x 4 (transcript type: control, inadmissible no-warning, inadmissible warning, and admissible) between subjects ANOVA was conducted to examine reported confidence levels. The confidence in the verdict rating was on a 1 (*not at all confident*) to 7 (*very confident*) scale. The means and standard deviations for confidence level as a function of the two factors appear in Table 2. There was no main effect for transcript type, no main effect for the verdict, and there was no significant interaction between the factors.

Case Influences

The jurors answered two questions about cases presented by the prosecution and defense. Jurors assessed the quality of the presentation of the case for both the prosecution and the defense on a 7-point scale (ranging from *not well* to *well*), and the strength of the evidence in

TABLE 1

Verdict Counts by Condition

| Verdict | Conditions | | | |
|------------|------------|-------------------------|----------------------|------------|
| | Control | Inadmissible No-Warning | Inadmissible Warning | Admissible |
| Guilty | 6 (25%) | 11 (50%) | 8 (34.8%) | 12 (52.2%) |
| Not Guilty | 18 (75%) | 11 (50%) | 15 (65.2%) | 11 (47.8%) |

TABLE 2

Means and Standard Deviations of Confidence in Verdict

| Conditions | Control | | Inadmissible No-Warning | | Inadmissible Warning | | Admissible | |
|------------|----------|-----------|-------------------------|-----------|----------------------|-----------|------------|-----------|
| | <i>M</i> | <i>SD</i> | <i>M</i> | <i>SD</i> | <i>M</i> | <i>SD</i> | <i>M</i> | <i>SD</i> |
| Guilty | 5.17 | 1.17 | 5.27 | 1.01 | 4.63 | .92 | 4.58 | 1.68 |
| Not Guilty | 4.28 | 1.53 | 4.36 | 1.12 | 5.13 | 1.41 | 4.27 | 1.19 |
| Total | 4.50 | 1.48 | 4.82 | 1.14 | 4.96 | 1.26 | 4.43 | 1.44 |

the case for both the prosecution and the defense on a separate 7-point scale (ranging from *not strong* to *strong*). The sum of these indicators demonstrated an overall opinion of the case presented by each side. Then, the subtraction of the rating of the case presented by the defense from the rating of the case presented by the prosecution measured which case the jurors found more convincing. Two internal consistency estimates of reliability determined the reliability of the defense and prosecution questions. Values for the coefficient alpha were both satisfactory at .85 for the prosecution responses and .76 for the defense responses. A one-way ANOVA was conducted on the cumulative scores, which did not differ among the four conditions. The means and standard deviations for the cumulative scores as a function of the condition appear in Table 3.

Next, a test of consistency analyzed the jurors' ratings of the cases presented with their verdicts. An independent samples *t* test analyzed the totals from this measure of defense minus the prosecution case strength responses across the verdicts. Those giving a guilty verdict reported the prosecution had a stronger case ($M = -2.73$, $SD = 3.15$) and those giving a not guilty verdict reported the defense had a stronger case ($M = 1.98$, $SD = 2.24$), $t(90) = 8.01$, $p < .001$. This test verifies that the participants were considering the case evidence when giving their verdict.

A one-way ANOVA examined the difference between the prosecution's ethicality score and that of the defense across the four experimental conditions.

The results yielded a significant difference $F(3, 88) = 2.73$, $p = .049$. The means and standard deviations for the ethicality score as a function of the condition appear in Table 3. The jurors found the prosecution less ethical than the defense in the warning condition compared to the admissible condition. None of the other differences were significant.

Discussion

Contrary to the stated hypotheses, there were no significant differences in the verdicts of the jurors among the four conditions. This finding suggests that the jurors were able to successfully disregard the inadmissible evidence, although it appears they also disregarded the evidence when it was admissible. Thus, these conditions failed to find the expected backfire effect or any effect of the inadmissible evidence. The findings are slightly disconcerting due to the fact that the admissible evidence condition did not yield higher conviction rates than the control condition. In this condition, the jurors read uncontradicted wiretap evidence of a confession to the crime, yet they did not convict in significantly higher numbers than the control. The result may be an indication of a couple of variables that affected the results of the experiment. The first is that despite the design of the experiment the participants simply did not read the transcript presented to them. This is a possibility, but the data showing that the jurors rated the defense's case stronger when they acquitted the defendant and the prosecution's case stronger when they convicted the defendants suggest that the participants were attentive to the trial transcript and the questions they answered about their impressions of the trial.

There may also be an explanation more consistent with theories of the backfire effect. First, these findings indicate that jurors do not trust wiretap evidence because of some larger societal distrust of this type of evidence and the way it may be obtained. This finding would accord with the discounting and suspicion theory. The jurors may have found the wiretap evidence inherently suspicious. The evidence may be so strong that the participants may have presumed unethical police tactics in obtaining the evidence. Even though the introduction of a warning made no difference in the verdicts, the findings of this experiment are not contrary to Schul (1993). There seemed to be a high suspicion of the wiretap evidence in all the conditions and the jurors disregarded that evidence in all conditions. Thus, the experiment may have been insufficient to test Schul's hypothesis.

However, the findings of this study may also support the pursuit of justice theory. It could be that the jurors in their own pursuit of justice did not believe the wiretap evidence to be reliable. Again, this could be because

TABLE 3

Means and Standard Deviations by Condition for Ratings of Effectiveness and Ethicality of Prosecution and Defense Cases

| Conditions | Effectiveness of Defense Minus Effectiveness of Prosecution | | Ethicality of Defense Minus Ethicality of Prosecution | |
|-------------------------|---|-----------|---|-----------|
| | <i>M</i> | <i>SD</i> | <i>M</i> | <i>SD</i> |
| Control | .54 | 3.36 | -.29 | 1.71 |
| Inadmissible No-Warning | -1.18 | 4.91 | .32 | 1.96 |
| Inadmissible Warning | .91 | 3.32 | -.96 | 1.61 |
| Admissible | -.13 | 1.98 | .48 | 2.23 |

the evidence was so strong that they thought there was no possible way it could be reliable. It is possible that jurors may pursue justice despite the evidence.

The inadmissible evidence did not affect the confidence of the participants in their verdict. The inadmissible evidence could have affected the confidence level in that the participants would have thought the prosecution to be less ethical when presenting inadmissible evidence, making them more uncertain about the validity of the other evidence and less confident in their verdict. However, no differences were found between most of the conditions in the measures of ethicality, in which participants rated how ethical they thought the defense and prosecution were during the trial on a 1 to 7 scale, or in the confidence level measurement. This suggests that the jurors completely disregarded the inadmissible evidence. The only difference among the conditions in ethicality showed that the jurors found the prosecution to be less ethical in the warning condition and the defense to be less ethical in the admissible condition. This finding shows that the judge's admonition combined with warnings may have played a role in reducing guilty verdicts by making the jurors suspicious of the prosecution and the prosecution's evidence, but the experiment's low number of participants may have hindered a finding of statistical significance.

The use of a capital case for a trial transcript may have affected the results of the experiment. During a capital case jurors are dismissed if they are not willing to impose the death penalty during the voir dire process, which involves questioning of the jurors for biases at the beginning of the trial. The dismissal of participants against the penalty in line with practice in capital cases did not occur, which could slightly skew the results. Jurors against the death penalty are more likely to acquit in a capital case. These jurors may also be more likely to disregard the inadmissible evidence. Including these jurors in the experiment made the jury inherently less likely to convict and could partially account for the lower conviction rates.

Another limitation of the current research is that it may only apply to inadmissible evidence that is actually unreliable. The admonishment by the judge states that the evidence was deemed inadmissible specifically because it is unreliable. If the evidence was deemed inadmissible because it was illegally obtained then the jurors may not have disregarded it. Sommers and Kassin (2001) demonstrated this result in previous research.

This study has some implications for psychology and law. It may add to the body of research that suggests inadmissible evidence can be disregarded by jurors completely if the proper procedures are implemented. In summary, this study neither confirms nor denies the pursuit of justice theory or discounting and suspicion theory as an effective solution to the problem of the backfire effect. The backfire effect permeates and corrupts the integrity of our court system, but more experimentation is needed to find the actual cause and the most effective remedy.

References

- American Psychological Association (2002). Ethical principles of psychologists and code of conduct. *American Psychologist, 57*, 1060-1073.
- Brehm, J. W. (1966). *A theory of psychological reactance*. New York: Academic Press.
- Brewer, N., & Williams, K. D. (Eds.). (2005). *Psychology and law: an empirical perspective*. New York: The Guilford Press.
- Cook, A., Arndt, J., & Lieberman, J. D. (2004). Firing back at the backfire effect: The influence of mortality salience and nullification beliefs on reactions to inadmissible evidence. *Law and Human Behavior, 28*, 389-410.
- Fein, S., McCloskey, A. L., & Tomlinson, T. M. (1997). Can the jury disregard that information? The use of suspicion to reduce the prejudicial effects of pretrial publicity and inadmissible testimony. *Personality and Social Psychology Bulletin, 23*, 1215-1226.
- Kassin, S. M., & Sommers, R. (1997). Inadmissible testimony, instructions to disregard, and the jury: Substantive versus procedural considerations. *Personality and Social Psychology Bulletin, 23*, 1046-1054.
- Kassin, S. M., & Sukel, H. (1997). Coerced confessions and the jury: An experimental test of the "harmless error" rule. *Law and Human Behavior, 21*, 27-46.
- Lee, D. H., Krauss, D. A., & Lieberman, J. (2005). The effects of judicial admonitions on hearsay evidence. *International Journal of Law and Psychiatry, 28*, 589-603.
- London, K., & Nunez, N. (2000). The effect of jury deliberations on jurors' propensity to disregard inadmissible evidence. *Journal of Applied Psychology, 85*, 932-939.
- Pickel, K. L. (1995). Inducing jurors to disregard inadmissible evidence: A legal explanation does not help. *Law and Human Behavior, 19*, 407-424.
- Schul, Y. (1993). When warning succeeds: The effect of warning on success in ignoring invalid information. *Journal of Experimental Social Psychology, 29*, 42-62.
- Sommers, S. R., & Kassin, M. (2001). On the many impacts of inadmissible testimony: Selective compliance, need for cognition, and the overcorrection bias. *Personality and Social Psychology Bulletin, 27*, 1368-1377.
- Stebly, N., Hosch, H. M., Culhane, S. E., & McWethy, A. (2006). The impact on juror verdicts of judicial instruction to disregard inadmissible evidence: A meta-analysis. *Law and Human Behavior, 30*, 469-492.
- Wissler, R. L., & Saks, J. (1985). On the inefficacy of limiting instructions: When jurors use prior conviction evidence to decide on guilt. *Law and Human Behavior, 9*, 37-48.

APPENDIX

The following includes the additional evidence and instructions presented in the experimental conditions.

Inadmissible Condition

Wire Tap

- MR KERNING: Was any evidence obtained in the investigation following the murders?
- MR HEFFLING: Yes, the defendant placed a phone call minutes after leaving the scene of the crime to a friend whose phone was wire tapped due to an unrelated case. According to the transcript of that conversation the defendant can be heard saying, "I killed Marylou and some bastard she was with. God, I don't... yeah, I ditched the blade".
- MR EVANS: Objection.
- THE COURT: Sustained. I would ask the jury to disregard the tape conversation as the evidence has been deemed inadmissible. It is inadmissible because the tape was barely audible and it was too difficult to determine what was said.

Warning Condition

Extra Instructions

- THE COURT: I would like to forewarn the jury that there may be evidence presented here today that has been deemed inadmissible by the court, because it was unreliable. If this should occur I will instruct you to disregard the evidence and it should have no effect on your verdict.

Wire Tap

- MR KERNING: Was any evidence obtained in the investigation following the murders?
- MR HEFFLING: Yes, the defendant placed a phone call minutes after leaving the scene of the crime to a friend whose phone was wire tapped due to an unrelated case. According to the transcript of that conversation the defendant can be heard saying, "I killed Marylou and some bastard she was with. God, I don't... yeah, I ditched the blade".
- MR EVANS: Objection.
- THE COURT: Sustained. I would ask the jury to disregard the tape conversation as the evidence has been deemed inadmissible. It is inadmissible because the tape was barely audible and it was too difficult to determine what was said. As I stated in the opening instructions inadmissible evidence is to be completely disregarded by the Jury, because this evidence has been determined to be unreliable.
-

Strong Body, Strong Mind: Self-Efficacy as a Function of Exercise

MARY BETH MILLER
Westminster College

This study examined how exercise affected the general self-efficacy of college students. Eighteen nonathlete undergraduates, 7 men and 11 women, between 18 and 22 years of age exercised for 2 weeks and then did not exercise for 2 weeks. Congruent with previous research, exercise significantly decreased stress levels and increased exercising self-efficacy of participants in comparison to their baseline scores. Overall, general self-efficacy scores significantly increased with exercise, the effect being significantly more pronounced in nonhabitual than in habitual exercisers. In contrast to past research, initial amount of stress did not significantly affect gains in either exercising or general self-efficacy. Collectively, the results suggest that exercising can build people's beliefs in their abilities.

“I can usually handle whatever comes my way,” though it is not comprehensive, this attitude reflects Jerusalem and Schwarzer’s (1979) definition of the psychological construct of self-efficacy, one’s perception of how capable of success one is in specific situations. Such perceptions are highly malleable through mastery experience, modeling, verbal persuasion, and physiological states (Bandura, 1997). The plethora of research on self-efficacy initiated by Bandura has shown that self-efficacy significantly affects persistence and accuracy in mathematical tasks (Bandura & Schunk, 1981; Schunk, 1981) and may affect young people’s decisions about their futures (Betz & Hackett, 1981). Such findings solidify the importance of maintaining a high sense of self-efficacy across one’s life.

Self-efficacy is, in a sense, self-confidence in specific situations (Bandura, 1997). People anticipate outcomes of a situation dependent on how good they judge their performance will be in that situation. A young woman may have low exercising self-efficacy, for example, because she does not believe that she can run at a moderate intensity for 30 min. Yet the same young woman may have high academic self-efficacy because she is completely certain that she can maintain a 4.0 throughout her college career. One’s general self-efficacy, then, is the accumulation of all these perceptions of one’s abilities in specific situations. It is one’s judgment of one’s ability to accomplish goals and solve

problems on a daily basis.

Self-efficacy is negatively correlated with psychological strain (Bandura, Cioffi, Taylor, & Brouillard, 1988; Jerome et al., 2002; Jex, Bliese, Buzzell, & Primeau, 2001). The ability to control cognitive demands reduces stress and mental strain and allows for normal cognitive functioning, and, according to Wiedenfeld et al. (1990), mastery of stressors strengthens one’s self-efficacy as well. In their study, Wiedenfeld et al. (1990) showed that as people with a phobia of snakes neared the day of the efficacy-acquisition phase, on which they would be expected to handle a snake, their stress levels rose significantly; but, as they gained experience in coping with the snake, their perceived self-efficacy increased. Cortisol levels were elevated at the beginning of each phase but declined as the sessions progressed, reaching their highest elevation in anticipation of the maximal-self-efficacy session. Once participants were fully efficacious, however, cortisol levels remained low, suggesting that mastery of stressors not only strengthens one’s sense of self-efficacy but also creates lasting changes that can protect against adverse immunological effects of stress.

The physiological effects of stress are measured easily, and research has found that exercise directly counteracts them. For example, King, Taylor, and Haskell (1993) found that their participants who were

Faculty mentor: Ted Jaeger

assigned to an exercise group reported significantly lower stress during a 12-month period than those in the control group. Similarly, Chafin, Christenfeld, and Gerin (2008) found that exercising after stress accelerated cardiovascular recovery in undergraduate students. Although the blood pressures and heart rates of those in the stress/exercise group were more elevated immediately after the exercise session than those of the participants in the other conditions, 10 min later (after the recovery period), their blood pressures were significantly closer to baseline than those of the participants who did not exercise after stress.

Though exercise counteracts the physiological effects of stress, its benefits are modulated by individual and group factors such as gender and fitness. Among men between 20 and 30 years of age, for example, those who had been trained in an aerobic fitness program had significantly lower heart rates than those who were untrained. In addition, although all participants displayed marked elevations in heart rate and arousal at the onset of stress, the trained group had significantly lower anxiety scores post-stress and their heart rates returned to baseline much more rapidly than those who were untrained (Sinyor, Schwartz, Peronnet, Brisson, & Seraganian, 1983). A similar study (Wilfley & Kuncze, 1986) revealed that exercise was much more beneficial to participants who were initially more stressed and had below-average fitness levels.

To determine whether exercise influences self-efficacy as it does stress, McAuley, Courneya, and Lettunich (1991) studied the effects of acute and long-term exercise on the perceived exercising self-efficacy of sedentary, middle-aged adults. Exercising self-efficacy was operationalized as the participant's belief in his or her ability to exercise for a certain amount of time without stopping. Participants displayed lowered resting heart rate and increased aerobic capacity at the end of the 20-week exercise program. More importantly, all participants indicated they had significantly higher exercising self-efficacy. More recently, McAuley, Elavsky, Jerome, Konopack, and Marquez (2005) found that physical activity itself, rather than actual physical fitness, may affect exercising self-efficacy. When 174 sedentary adults were separated into high-efficacy and low-efficacy conditions (manipulated by providing bogus feedback regarding initial performance), the high exercising self-efficacy group reported greater positive well-being than their less efficacious counterparts at baseline, 2 months, and 4 months. At 6 months, however, the perceived well-being of the high efficacy group had been surpassed by that of the low efficacy group, whose reports of positive affect increased linearly with time. Thus, the authors suggest that the psychological benefits of exercise may be more strongly related to physical activity than to

exercise ability or physical fitness.

While the aforementioned studies (McAuley et al., 1991; McAuley et al., 2005) have established a specific connection between exercise and increases in *exercising* self-efficacy, the effect of aerobic exercise on general self-efficacy is unexplored. This relationship would be particularly important for college students because of the finding that *general* self-efficacy can influence people's decisions about their futures (Betz & Hackett, 1981).

Based on the findings of King et al. (1993) and McAuley et al. (1991), I hypothesized that long-term, aerobic exercise would increase the general and exercising self-efficacy of a sample of college students. Also, given Sinyor et al.'s (1983) results, I hypothesized that habitual exercisers would experience less of an increase in general and exercising self-efficacy following regular exercise than would nonhabitual exercisers. Finally, based on the findings of Wilfley and Kuncze (1986), I predicted that less perceived stress would yield smaller increases in general and exercising self-efficacy.

Method

Participants

After receiving an email inviting them to participate in an exercise study, 27 undergraduates who were not currently college athletes agreed to work out for 2 weeks and then not to work out for 2 weeks as part of an experiment. Of the 18 students who successfully completed the 4-week study, 7 were men and 11 were women, and their ages ranged from 18 to 22 years. The 12 participants who were also psychology students received extra credit in their psychology courses, and, as additional incentive, the names of all participants who successfully completed the study were placed in a raffle for prizes of \$25, \$75, and \$100 cash.

Materials

Workout log. Each participant was supplied a workout log on which to record the day of the workout, the type of workout he or she chose to do, and the length of the workout.

Self-efficacy measures. Participants filled out the 10-item General Self-Efficacy Scale (Jerusalem & Schwarzer, 1979). They rated on a 4-point scale their agreement with statements such as "I can always manage to solve difficult problems if I try hard enough" and "If someone opposes me, I can find the means and ways to get what I want." General self-efficacy scores, then, ranged from 0 to 40, with higher scores indicating higher levels of general self-efficacy. The unidimensionality of the scale has been confirmed by studies in 14 countries (Schwarzer & Scholz, 2000). An exercising self-efficacy scale (Hu, Motl, McAuley, &

Konopack, 2007) measured how capable participants believed they were of completing 10, 20, 30, 40, 50, and 60 min of aerobic exercise at a moderate intensity. Participants rated how confident they were in their ability to successfully carry out the specified activity for each item on a scale from 0 to 100 presented in 10-point increments (a score of 100 meant they were 100 percent confident they could run for the specified amount of time).

Stress measure. Participants measured their stress using the 14-item Perceived Stress Scale (Cohen, Kamarck, & Mermelstein, 1983). They rated how often in the last month they had experienced sensations such as being “upset because of something that happened unexpectedly” and feeling “that you were unable to control the important things in your life” on a scale ranging from 1 (*never*) to 5 (*very often*). Total scores ranged from 0 to 70, with higher scores indicating more perceived stress. This scale showed reliability as tested by Cohen, et al. (1983).

Procedure

Participants met in the intramural gym in order to receive oral instructions and to sign their consent forms. After being separated into two groups (habitual exercisers, who reported already working out more than three days in an average week, and nonhabitual exercisers, who reported not working out regularly), participants were randomly assigned to either Group A, whose participants would work out only during the first 2 weeks of the experiment (phase one), or to Group B, whose participants would work out only during the second 2 weeks of the experiment (phase two). The two phases of the experiment were implemented to control for stress induced by regular cycles of testing.

Working out consisted of walking, running, swimming, riding an exercise bike, or using an elliptical machine. Participants were told that the exercise needed to be continuous but it did not need to be 30 min at the same stress level (they were allowed to walk in-between periods of running, for example, but they were not allowed to *stop* walking). They were told to work out as long as they would like, but 30 min was the minimum for participation in the study. It was made clear that they were only to work out during the appropriate phase of the study. During nonactive weeks, they were not to engage in any physical activity geared toward becoming more physically fit. If they did or did not work out during the appropriate phase, however, they were asked to report honestly in their workout logs.

All participants were assured of the confidentiality of the study, informed of their right to stop partici-

pation at any time without penalty, and invited to ask questions. Participants then completed the Perceived Stress Scale, the General Self-Efficacy Scale, and an exercising self-efficacy scale; then the researcher thanked them for their participation. Thereafter, participants met with the researcher on either Monday or Tuesday of each week for the next 4 weeks to fill out the appropriate measures. They completed the general self-efficacy and exercising self-efficacy scales prior to beginning the experiment, after phase one, and after phase two. They completed the perceived stress scale prior to beginning the experiment and every week thereafter. Participants were debriefed upon completion of the month-long study.

Results

We calculated the general self-efficacy, exercise efficacy, and stress scores for each participant. After reverse-scoring the appropriate perceived stress items, we computed perceived stress and exercise efficacy scores by summing participant responses to the individual items of both measures. Exercise self-efficacy scores were based on 0 to 100 percent confidence; total scores were summed and then divided by 100 to obtain an average percent confidence score. On the general self-efficacy scales, responses of “not true at all,” “hardly true,” “moderately true,” and “exactly true” were assigned numbers of 1 through 4, respectively. Responses to all items were then summed to determine each participant’s general self-efficacy score, which ranged from 0 to 40. Final scores of each participant’s measures were then transformed into scores that would reflect the way in which participants’ general self-efficacy, exercising self-efficacy, and stress scores had changed over the course of the study. For those who had worked out during phase one of the experiment, their baseline general self-efficacy, exercising self-efficacy, and stress scores were subtracted from their general self-efficacy, exercising self-efficacy, and stress scores after 2 weeks of exercise. For those who had exercised during phase two of the experiment, their general self-efficacy, exercising self-efficacy, and stress scores after 2 weeks of not working out were subtracted from their general self-efficacy, exercising self-efficacy, and stress scores after 2 weeks of exercise.

Change scores were analyzed using a two-way mixed model ANOVA, the two independent variables being type of exerciser (habitual or nonhabitual) and phase of exercise (exercising or not exercising), with repeated measures on the phase of exercise variable. As suggested in Figure 1, general self-efficacy was significantly affected by phase of exercise, $F(1,14) = 7.05$, $p = .017$, partial $\eta^2 = .306$, meaning that general

self-efficacy scores of both habitual and nonhabitual exercisers increased significantly when working out for 2 weeks ($M = 1.78, SD = 2.46$) as compared to their scores when not allowed to work out for 2 weeks ($M = -1.11, SD = 3.03$). Type of exerciser also had a significant impact on general self-efficacy scores, $F(1, 14) = 4.50, p = .05, \eta^2 = .219$, in that nonhabitual exercisers reported more positive change in general self-efficacy regardless of phase of exercise.

The data presented in Figure 1 also show a similar and significant effect of exercising on participants' exercising self-efficacy, $F(1,14) = 7.27, p = .016$, partial $\eta^2 = .312$. Specifically, exercising self-efficacy scores increased by a mean of 5.00 ($SD = 6.15$) during participants' 2 weeks of working out; while not working out, they remained relatively stable, changing only by a mean of 0.94 ($SD = 7.67$). Type of exerciser did not significantly affect exercising self-efficacy, although this effect did approach significance, $F(1,14) = 4.10, p = .06$, partial $\eta^2 = .204$.

As indicated by Figure 2, exercise also significantly affected perceived stress scores, $F(1,14) = 6.20, p = .024$, partial $\eta^2 = .280$, while type of exerciser did not. Participants reported an average 7.78 decrease ($SD = 9.42$) in stress when working out and an average increase of 2.17 ($SD = 9.97$) in stress when not allowed to work out.

Finally, to determine if less stress predicted smaller increases in general or exercising self-efficacy, participants with moderate stress scores (between 52 and 62) were removed from the sample and the remaining participant scores were divided into two groups: low stress (those with scores between 36 and 51) and high stress (those with scores between 63 and

78), resulting in 6 participants in each group. The mean increase in general self-efficacy and exercising self-efficacy was compared using independent samples t tests. Findings were not significant for general self-efficacy nor for exercising self-efficacy, indicating that level of stress did not have an impact on level of self-efficacy.

Discussion

The present study showed that aerobic exercise significantly increased the general self-efficacy of college students, regardless of their exercise history. That is, after 2 weeks of engaging in at least 30 min of aerobic exercise three times a week, participants reported significantly higher levels of general self-efficacy than they reported either at baseline or after 2 weeks of not exercising. However, when separated by type of exerciser, nonhabitual exercisers reported significantly more positive changes in general self-efficacy than did habitual exercisers. Participants also reported higher levels of exercising self-efficacy, as well significantly lower levels of perceived stress, than they did after 2 weeks of not exercising. Finally, level of perceived stress did not appear to have a significant impact on general or exercising self-efficacy.

The unique finding of the present study was the effect of aerobic exercise on the general, rather than exercising, self-efficacy of college students. Previous research (McAuley et al., 1991; McAuley et al., 2005) testing the effects of exercise on self-efficacy has focused solely on exercising self-efficacy. Thus, the finding of an increase in general self-efficacy is distinct-

FIGURE 1

Line graphs showing the mean change from baseline in general and exercising self-efficacy scores for both habitual and nonhabitual exercisers as related to phase of exercise (either exercising or not exercising)

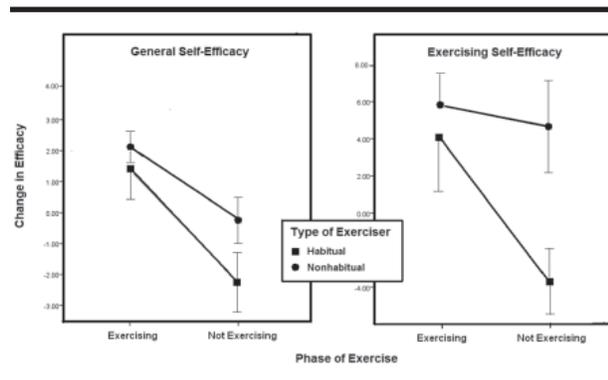
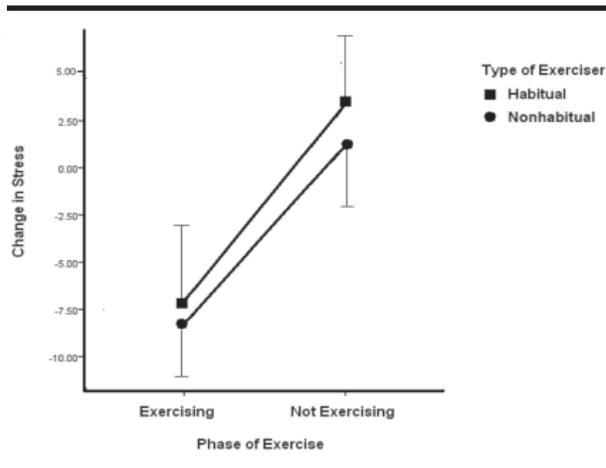


FIGURE 2

Line graph showing the mean change from baseline in stress scores for both habitual and nonhabitual exercisers as related to phase of exercise (either exercising or not exercising)



tive. Not surprisingly, nonhabitual exercisers reported essentially no change in general self-efficacy when not exercising. Their reports of general self-efficacy increased, however, when working out. Congruent with previous research (Wilfley & Kuncze, 1986), this effect was significantly more pronounced in nonhabitual versus habitual exercisers. Habitual exercisers reported significantly less positive changes in general self-efficacy than did nonhabitual exercisers. Though their general self-efficacy increased when exercising, the general self-efficacy of habitual exercisers decreased significantly when they were not allowed to exercise for 2 weeks. The finding that habitual exercisers were more affected by lack of exercise than were nonhabitual exercisers makes sense because not exercising violates part of a habitual exerciser's weekly routine. It is possible that these exercisers were intrinsically motivated to work out on a regular basis and that when not allowed to work out the lack of exercise negatively affected their psychological well-being.

Both groups of exercisers reported higher levels of exercising self-efficacy after 2 weeks of exercise as compared to baseline and after 2 weeks of nonexercise, as suggested by previous research (McAuley et al., 1991; McAuley et al., 2005). Among habitual exercisers, exercising efficacy increased while working out and decreased while not exercising. Interestingly enough, however, nonhabitual exercisers reported an increase in exercising efficacy even when they were not exercising. Though only marginally significant, this difference between groups is quite apparent in Figure 1 and should be further explored in a similar study with more participants.

Given previous data (Chafin et al., 2008; King et al., 1993; Wilfley & Kuncze, 1986), I expected that exercise would significantly decrease participants' perceived stress. It was surprising, though, that level of stress was not affected by an increase in either general or exercising self-efficacy. Perhaps a larger sample size would have revealed the expected relationships. It is important to note that the present study included a sample of only 18 participants because 9 of the 27 participants either dropped out or did not complete the experiment as instructed. Although the groups were divided fairly evenly (7 men, 11 women; 8 habitual exercisers, 10 nonhabitual exercisers), the small sample size may account for the lack of significant effects for level of stress.

The current study leaves several aspects of the relationship between exercise and self-efficacy unexplored. Future research could explore the exercise/self-efficacy relationship among athlete and nonathlete students. Athletes were not included in the present study because their enforced intensive

exercise may either make them immune to, or make them react especially strongly to, 2 weeks of not exercising. Also, despite 30 years of active research, gender differences in general self-efficacy are poorly understood. Because gender differences have not been well-defined by previous research, gender differences in general and exercising self-efficacy need further clarification. Finally, the finding that not exercising impacts the general self-efficacy of habitual exercisers has not been reported before and, therefore, warrants exploration. The general self-efficacy of habitual exercisers drops considerably when they are not exercising (see Figure 1). Future research might investigate the relationship between general self-efficacy and self-esteem (or, more specifically, self-esteem variability) and the way in which exercise affects the combination. It may be that people with more stable self-esteem will be less affected by exercise than those with more variable self-esteem (Kernis, 2005).

Regardless of limitations, the present study yielded findings that have important implications for the lives of college students. General self-efficacy significantly affects performance on a variety of tasks (Bandura & Schunk, 1981; Schunk, 1981) and may affect students' decisions about their life goals (Betz & Hackett, 1981). The knowledge that exercise can build one's beliefs in one's abilities should challenge young people to re-evaluate their lifestyles and to incorporate exercise into their lives.

References

- Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York, NY: W.H. Freeman Company.
- Bandura, A., Cioffi, D., Taylor, C. B., & Brouillard, M. E. (1988). Perceived self-efficacy in coping with cognitive stressors and opioid activation. *Journal of Personality and Social Psychology, 55*, 479-488.
- Bandura, A., & Schunk, D. H. (1981). Cultivating competence, self-efficacy, and intrinsic interest through proximal self-motivation [Abstract]. *Journal of Personality and Social Psychology, 41*, 586-598.
- Betz, N. E., & Hackett, G. (1981). The relationship of career-related self-efficacy expectations to perceived career options in college women and men [Abstract]. *Journal of Counseling Psychology, 28*, 399-410.
- Chafin, S., Christenfeld, N., & Gerin, W. (2008). Improving cardiovascular recovery from stress with brief post-stress exercise. *Health Psychology, 27*, S64-S72.
- Cohen, S., Kamarck, T., & Mermelstein, R. (1983). A global measure of perceived stress. *Journal of Health and Social Behavior, 24*, 385-396.
- Hu, L., Motl, R. W., McAuley, E., & Konopack, J. F. (2007). Effects of self-efficacy on physical activity enjoyment in college-aged women. *International Journal of Behavioral Medicine, 14*, 92-96.
- Jerome, G. J., Marquez, D. X., McAuley, E., Canaklisova, S., Snook, E., & Vickers, M. (2002). Self-efficacy effects on feeling states in women. *International Journal of Behavioral Medicine, 9*, 139-154.
- Jerusalem, M., & Schwarzer, R. (1979). *The general self-efficacy scale (GSE)*. Retrieved October 8, 2009, from <http://userpage.fu-berlin.de/health/engscal.htm>.

- Jex, S. M., Bliese, P. D., Buzzell, S., & Primeau, J. (2001). The impact of self-efficacy on stressor-strain relations coping style as an explanatory mechanism. *Journal of Applied Psychology, 86*, 401-409.
- Kernis, M. H. (2005). Measure self-esteem in context: The importance of stability of self-esteem in psychological functioning [Electronic version]. *Journal of Personality, 73*, 1569-1605.
- King, A. C., Taylor, C. B., & Haskell, W. L. (1993). Effects of differing intensities and formats of 12 months of exercise training on psychological outcomes in older adults. *Health Psychology, 12*, 292-300.
- McAuley, E., Courneya, K. S., & Lettunich, J. (1991). Effects of acute and long-term exercise on self-efficacy responses in sedentary, middle-aged men and women. *The Gerontologist, 31*, 534-542.
- McAuley, E., Elavsky, S., Jerome, G. J., Konopack, J. F., & Marquez, D. X. (2005). Physical activity-related well-being in older adults' social cognitive influences. *Psychology and Aging, 20*, 295-302.
- Schunk, D. H. (1981). Modeling and attributional effects on children's achievement: A self-efficacy analysis [Abstract]. *Journal of Educational Psychology, 73*, 93-105.
- Schwarzer, R., & Scholz, U. (2000). *Cross-cultural assessment of coping resources: The general perceived self-efficacy scale*. Paper presented at the First Asian Congress of Health Psychology: Health Psychology and Culture, Tokyo, Japan. Retrieved October 10, 2008, from <http://www.statisticssolutions.com/general-self-efficacy-schwarzer>.
- Sinyor, D., Schwartz, S. G., Peronnet, F., Brisson, G., & Seraganian, P. (1983). Aerobic fitness level and reactivity to psychosocial stress: Physiological, biochemical, and subjective measures. *Psychosomatic Medicine, 45*, 205-217.
- Wiedenfeld, S. A., O'Leary, A., Bandura, A., Brown, S., Levine, S., & Raska, K. (1990). Impact of perceived self-efficacy in coping with stressors on components of the immune system. *Journal of Personality and Social Psychology, 59*, 1082-1094.
- Wilfley, D., & Kuncze, J. (1986). Differential physical and psychological effects of exercise. *Journal of Counseling Psychology, 33*, 337-342.

An Eye for an Eye: Examining Personality-Based Predictors of Trait Vengefulness

SEAN E. WISNIESKI
Shepherd University

Vengeance is a form of aggression aimed at retaliating against another for some perceived transgression. Previous research suggests a strong basis for the prediction of vengefulness from individual personality traits. The present study examined unforgiveness, narcissism, self-esteem, locus of control, trait anger, self-control, age, and sex as possible predictors. Participants (N = 110) completed measures of trait anger, externality of personal control, subclinical narcissism, forgiveness, vengefulness, and global self-esteem. Only age and externality failed to contribute to the model's predictive power, which otherwise accounted for more than half the variance in vengefulness scores. The results suggest a strong basis for predicting vengefulness at the individual level independent of situational factors.

Social interactions sometimes give rise to interpersonal offense, whether in the form of a tasteless joke, a hurtful remark, or a betrayed trust. When an individual feels hurt by someone, feelings of unforgiveness often result. Unforgiveness refers to a constellation of generally negative emotions resulting from the perception of an interpersonal transgression; such emotions may include anger, bitterness, hatred, resentment, and rumination (Berry, Worthington, O'Connor, Parrott, & Wade, 2005). Different people attempt to resolve, reduce, or avoid these negative feelings in different ways (Worthington & Wade, 1999). Some may choose to forgive their transgressor, freely relinquishing those negative emotions and any claims they might have to restitution or justice for the harm they suffered (Exline, Baumeister, Bushman, Campbell, & Finkel, 2004). When people genuinely forgive their transgressors, the result is a motivational transformation away from retaliation and estrangement and toward goodwill, compassion, and mercy (McCullough, Worthington, & Rachal, 1997). By contrast, some may choose to seek vengeance. Vengeance is a form of retributive aggression in which the victim of some transgression responds in a manner that is usually aggressive, disproportionate, intense, highly personal, and temporally incontinent to the harm suffered, typically following a period of angry rumination (Stuckless & Goranson, 1992).

Although both forgiveness and vengeance have the

effect of reducing the negative feelings associated with unforgiveness (Worthington & Wade, 1999), there is abundant evidence that the two are unequal in most other respects. For instance, the literature indicates that vengeance is a highly malicious and antisocial behavior implicated in a number of violent crimes, including arson (O'Sullivan & Kelleher, 1987), homicide (Porporino, Doherty, & Sawatsky, 1987), and rape (Scully & Marolla, 1985). Conversely, the literature paints forgiveness as a prosocial behavior that confers both physical and mental health benefits on its practitioners. Forgiveness-based therapeutic interventions, such as that designed by Enright and the Human Development Study Group (1991), have resulted in significant long-term mental health improvements in participating clients who identified as suffering from recent emotional distress (e.g., Al-Mabuk, Enright, & Cardis, 1995).

Thus, forgiveness and vengefulness, although clearly related, are empirically distinct phenomena.

Author note. The author extends his sincerest thanks to Leah Shaffer, Professor Judy Spence, and Christen Wall for their immeasurable assistance in the collection of data. The author also thanks Dr. Larry Z. Daily of Shepherd University, the faculty sponsor for this paper, whose feedback and kind patience was nothing less than critical in its completion.

Correspondences concerning this article should be addressed to Sean E. Wisnieski at 28 W. Orndorff Dr, Brunswick, MD, 21716-1220. E-mail: swisni02@shepherd.edu.

Faculty mentor: Ashley Tempel

Moreover, Brown (2003) noted that although vengefulness, by definition, entails unforgiveness, unforgiveness does not necessarily lead to vengefulness; some unforgiving people may, for a variety of reasons, prefer simply to harbor a grudge or avoid their transgressor. What, then, makes some people more vengeful than others? Brown (2004) suggested the influence of certain personality traits and found that trait narcissism exerted just such an influence. Brown (2003) proposed that this influence might be attributable to narcissists' elevated self-efficacy and self-entitlement, which might lead narcissists to feel more able and willing to exact vengeance. The purpose of the present research was to investigate several other personality traits, including trait anger, externality of personal control, low self-control, high self-esteem, sex, and age, which might contribute significantly to the prediction of vengefulness.

One reason for these predictions is that prior research (e.g., Wilkowski & Robinson, 2007) has demonstrated that aggressive thoughts and feelings often result from failures in cognitive self-regulation. Wilkowski and Robinson (2007) likened this self-regulation to a kind of psychological home security system, in which the intrusion of unwanted negative thoughts results in the recruitment of limited cognitive resources to repel them. When such resources become depleted, affected individuals may momentarily lose the ability to inhibit such thoughts, feelings, and behaviors. In accordance with this theory, Stucke and Baumeister (2006) found that the depletion of self-control resources under experimental conditions resulted in an increased tendency toward aggression. It is possible that such aggression might manifest in the form of vengeance-seeking behaviors.

Thus, it is probable that any personality characteristic that taxes this limited self-control resource will result in a reduced ability to inhibit vengeful behaviors. That neuroticism, for instance, promotes vengefulness (McCullough, Bellah, Kilpatrick, & Johnson, 2001) is likely a function of its documented tendency to predispose people to experience more interpersonal stress (Gunthert, Cohen, & Armeli, 1999), which depletes self-control and permits the intrusion of hostile and vengeful thoughts. Likewise, Wilkowski and Robinson's (2008) finding that individuals high in trait anger suffer from a relative dearth of self-regulation dovetails neatly with other findings implicating trait anger in unforgiveness and vengeful rumination (e.g., Berry et al., 2005).

For these reasons, I hypothesized that vengefulness could be predicted from a combination of traits that promote emotional stress that, in turn, deplete self-control, removing one internal barrier to the exact

of vengeance. Narcissism is one such trait that earlier research has identified as a key predictor of vengefulness (e.g., Brown, 2004). Narcissists may be more vengeful because they tend to perceive transgressions more readily than nonnarcissists (McCullough, Emmons, Kilpatrick, & Mooney, 2003). Narcissists are also more globally aggressive than nonnarcissists, to an extent not captured by simple, healthy self-esteem (Bushman & Baumeister, 1998), which may also predispose them to vengeance-seeking. Another hypothesized predictor of vengefulness is externality of personal control, defined by Lefcourt (1976) as the belief that events in life are determined by luck, fate, chance, or the intervention of powerful others. Lefcourt related externality to learned helplessness, a contention supported by Benassi, Sweeney, and Dufour's (1988) meta-analysis linking externality to depression. Because positive affect can help guard against the depletion of self-control (Tice, Baumeister, Shmueli, & Muraven, 2007), it logically follows that persistent states of negative affect, such as depression, would exacerbate the depletion of self-control and dispose affected individuals to more aggressive thoughts and behaviors. Moreover, Bayse, Allgood, and Van Wyk (1992), in a study of family life education in correctional rehabilitation, found that nearly half of the inmates sampled were significantly more external than average. No study has yet linked locus of control with vengefulness but, as with narcissism, the general disposition of external individuals toward negative affectivity and criminal aggression indicates a poor capacity for self-regulation. In addition, Muraven's (2008) finding that autonomous exercise of self-control tends to deplete the resource less quickly than when the exercise of self-control is compelled conforms with Baumeister, Vohs, and Tice's (2007) earlier claim that motivations and perceptions of self-control exercise can temporarily guard against the negative effects of its depletion. Because externality is characterized by feelings of helplessness and captivity to the caprices of fate or powerful others, externality likely taxes the inhibitory self-control resource and therefore promotes vengefulness. I also anticipated that trait anger would contribute to the prediction of vengefulness, because previous research has established a direct link between trait anger and vengefulness. Berry et al. (2005), for instance, confirmed that high trait anger both inhibited forgiveness and promoted vengeful rumination.

Baseline self-control will likely also help predict vengefulness. Although all individuals possess self-control to some extent, there is ample evidence that this capacity varies from individual to individual (Tangney, Baumeister, & Boone, 2004). Accordingly, individuals with low trait self-control are likely less capable of inhibiting vengeful impulses, even in the

absence of traits that might tax self-control extensively, such as narcissism. This notion is critical because prior research (Muraven, Pogarsky, & Shmueli, 2006; Stucke & Baumeister, 2006) has implicated low self-control in retaliatory, violent aggression.

Finally, I hypothesized that age and sex would also relate to vengefulness. At least two studies (Cota-McKinley, Woody, & Bell, 2001; Stuckless & Goranson, 1992) have found that age correlated negatively with vengefulness and that men endorsed vengeful attitudes more than women.

Methods

Participants

One hundred and ten participants (57 women, 53 men) were sampled. Of these, 47 (42.7%) were Shepherd University students of various majors and area residents who participated on a voluntary basis, 41 (37.3%) were students enrolled in an introductory psychology course whose participation satisfied a course research participation requirement, and 22 (20.0%) were students in an introductory English course who participated for a small amount of extra credit. The mean age of the participants was 21.90 ($SD = 6.05$). The recruitment of human participants was approved by Shepherd University's Institutional Review Board prior to the start of the study, and all participants signed informed consent forms prior to participation.

Measures

I used seven inventories to assess forgivingness, vengefulness, trait anger, narcissism, self-esteem, self-control, and locus of control. Participants also completed a brief demographic questionnaire to record their age, sex, and reason for participation.

Forgivingness. I assessed forgivingness using Brown's (2003) Tendency to Forgive (TTF) Scale, a four-item measure intended to assess individual differences in forgivingness. Despite its brevity, the TTF is adequately reliable, with Cronbach's internal consistency coefficients in the .80 to .90 range and test-retest reliability of .71 (Brown, 2003; Brown & Phillips, 2005). The measure yields a final score ranging from 1.00 to 7.00, with higher scores representing higher forgivingness. Sample items include "I tend to get over it quickly when someone hurts my feelings" and "I have a tendency to harbor grudges" (reverse-scored).

Vengefulness. I measured vengefulness using the 20-item Vengeance Scale (VS) developed by Stuckless and Goranson (1992). Previous researchers have demonstrated the measure's reliability, finding internal consistency and test-retest reliability coefficients typically exceeding .80 (Brown, 2004; Holbrook, White, & Hutt, 1995; Hutt, Iverson, Bass, & Gayton, 1997).

The measure's final score ranges from 20 to 140, with higher scores representing a stronger endorsement of vengeful attitudes. Sample items include "There is nothing wrong in getting back at someone who has hurt you" and "It's not worth my time or effort to pay back someone who has wronged me" (reverse-scored).

Narcissism. I assessed participants' narcissism using a truncated version of the Narcissistic Personality Inventory (NPI). The NPI was originally developed by Raskin and Hall (1979) and consists of 50 items, with participants selecting which of a pair of sentences (one narcissistic, one nonnarcissistic) described them best. Raskin and Hall (1981) found that the NPI exhibited both adequate split-half and internal consistency reliability (both over .70). Emmons (1987) retained only 37 of the original 50 items, a true-false version of which I administered in this study per Brown (2004). I coded *true* responses as 1 and *false* responses as 0. I summed these responses to yield a final score between 0 and 37, with higher scores reflecting more extreme levels of narcissism. Sample items include "I usually dominate any conversation" and "I insist upon getting the respect that is due me."

Self-esteem. Several studies have established the importance of distinguishing narcissism from simple, healthy self-esteem (Brown & Zeigler-Hill, 2004; Bushman & Baumeister, 1998). To that end, I used Tafarodi and Swann's (2001) Revised Self-Liking (SL) Scale to assess self-esteem. The scale consistently exhibits internal consistency coefficients in the .90 and above range (Brown & Zeigler-Hill, 2004; Tafarodi & Swann, 2001). The SL produces a final score ranging from 8 to 40, with higher scores reflecting higher self-esteem. Sample items include "I never doubt my personal worth" and "I do not have enough respect for myself" (reverse-scored).

Trait anger. I used the Multidimensional Anger Inventory (MAI) to assess trait anger. Siegel (1986) developed the MAI in order to measure dimensions of anger that are clinically relevant for diagnoses of cardiovascular disease, but other researchers (e.g., Rogge & Bradbury, 1999) have established its usefulness as a measure of global trait anger, as well. Kroner, Reddon, and Serin (1992) found a full-scale internal consistency of .93 and factor internal consistencies of .90 and greater. The MAI consists of 38 statements that, when summed, yield a final score with a range from 38 to 190, with higher scores reflecting greater trait anger.

Locus of control. I used Duttweiler's (1984) Internal Control Index (ICI) to assess locus of control. The ICI is a measure of internality of personal control consisting of 28 statements, which I summed

to yield a final score ranging from 28 (*very external*) to 140 (*very internal*). At least two studies (Duttweiler, 1984; Meyers & Wong, 1988) have found Cronbach’s coefficients of .85, suggesting the measure is internally consistent. Sample items include “I like jobs where I can make decisions and be responsible for my own work” and “When faced with a problem, I try to forget it” (reverse-scored).

Self-control. In order to measure participants’ baseline self-control, I used the full version of the Self-Control (SC) scale developed by Tangney et al. (2004). The SC is a 36-item measure that yields a final score ranging from 36 to 180, with higher scores reflecting a greater capacity for self-control. Tangney et al. exhaustively established the measure’s convergent validity by measuring its relation to various traits such as depression, anxiety, hostility, and conscientiousness. Duckworth and Seligman (2006) found internal consistency of .86 for the brief version of the scale, and Tangney et al. reported a coefficient of .89 for the full version of the scale. Tangney et al. also found test-retest reliability of .89 and .87 for the full and brief versions of the scale, respectively. Sample items include “People

can count on me to keep on schedule” and “I have trouble saying no” (reverse-scored).

Procedure

I assembled the measures in packets in the following randomly determined order: NPI, TTF, VS, SL, ICI, MAI, and SC, with the brief demographic survey at the end. I informed participants that they were participating in a study of personality-based predictors of vengefulness. After briefing and informed consent, all participants completed their packets. Following completion of the study, I debriefed all participants and answered any questions. I analyzed the data using the Statistical Package for the Social Sciences (SPSS), version 17.0.

Results

Descriptive Statistics

Table 1 shows means, standard errors, minimum and maximum observed scores, and Cronbach’s α internal consistency coefficients for each of the measures used in the study. Because men scored significantly higher than women on both the Self-Control Scale,

TABLE 1

Descriptive Statistics for Measures

| Measure | <i>M</i> | <i>SE</i> | Min | Max. | α |
|---------|----------|-----------|------|------|----------|
| VS | 59.64 | 1.87 | 21 | 121 | .92 |
| TTF | 3.98 | 0.12 | 1.00 | 6.75 | .78 |
| ICI | 103.95 | 1.15 | 60 | 131 | .82 |
| MAI | 107.04 | 1.96 | 63 | 171 | .89 |
| NPI | 17.42 | 0.59 | 2 | 32 | .83 |
| SC | 113.69 | 1.72 | 64 | 157 | .89 |
| Men | 117.36 | 2.36 | 82 | 157 | -- |
| Women | 110.28 | 2.43 | 64 | 150 | -- |
| SL | 27.94 | 0.74 | 8 | 40 | .94 |
| Men | 29.79 | 0.90 | 13 | 40 | -- |
| Women | 26.21 | 1.12 | 8 | 40 | -- |

Note. VS = Vengeance Scale; TTF = Tendency to Forgive Scale; ICI = Internal Control Index;

MAI = Multidimensional Anger Inventory; NPI = Narcissistic Personality Inventory; SC = Self-Control Scale; SL = Revised Self-Liking Scale.

$t(108) = -2.08, p = .04, d = .04$, and the Self-Liking Scale, $t(104.42) = -2.49, p = .01, d = .05$, separate descriptive statistics are provided for men and women.

Zero-Order Correlations

Table 2 displays a correlation matrix for all measures used in the study.

Regression Analysis

To test the ability of the hypothesized personality traits to predict vengefulness, I regressed scores on the VS in a simultaneous multiple regression analysis as a function of scores on the TTF, ICI, MAI, NPI, SC, and SL scales, as well as sex (dummy-coded 1 for men and 0 for women) and age. Prior to conducting the analysis, I mean-centered (converted to deviation-score form) each predictor in order to promote interpretability of the results (Aiken & West, 1991). This regression analysis returned an adjusted R^2 of .53 ($SEE = 13.38, f^2 = 1.31$), which differed significantly from 0, $F(8, 101) = 16.62, p < .001, MS_E = 178.97$. Of the terms included in the equation, the ICI term, $\beta = -.02, t(101) = -0.25, p = .80$, and the age term, $\beta = -.04, t(101) = -0.54, p = .59$, were nonsignificant. In addition, the sex term did not reach significance, $\beta = 5.37, t(101) = 1.92, p = .06$.

Because the ICI and age terms were nonsignificant, I removed them from the analysis and recalculated the regression equation to remove bias from the remaining

partial slope coefficients (Berry & Feldman, 1985). I retained the sex term because it was near significance and its inclusion was strongly supported by existing theory. The revised regression analysis showed a negligible improvement in the model's goodness-of-fit, with an adjusted R^2 of .54 ($SEE = 13.27, f^2 = 1.31$), which also differed significantly from 0, $F(6, 103) = 22.44, p < .001, MS_E = 176.21$. Table 3 displays a regression analysis summary with standardized and unstandardized partial slope coefficients, standard errors, t -test statistics, and semipartial correlations with the criterion for each term.

Variance inflation factors, calculated on uncentered predictors per Belsley's (1984) recommendation, confirmed that the model was not contaminated by multicollinearity (all VIFs < 2.1). In addition, both examination of the residuals plot and the calculation of a White (1980) test statistic showed that the model was not heteroskedastic, $LM = 22.66, df = 26, p = .65$.

As a secondary analysis, I examined the mediating effect of trait anger on the relation between age and vengefulness. Calculation of a Goodman test statistic revealed that trait anger completely mediated the predictive power of age on vengefulness, $z = -1.97, p < .05$, with the partial correlation between age and vengefulness after controlling for trait anger dropping to nonsignificance, $pr(107) = -.14, p = .15$.

TABLE 2

Zero-Order Correlation Matrix

| Measure | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---------|---|-------------------|------|-------------------|------------------|-------------------|-------------------|-------------------|
| 1. VS | — | -.47 [†] | -.03 | .59 [†] | .48 [†] | .38 [†] | .11 | -.22 [*] |
| 2. TTF | — | — | .05 | -.60 [†] | -.10 | .20 [*] | .24 [*] | .08 |
| 3. ICI | — | — | — | -.10 | .24 [†] | .37 [†] | .35 [†] | .19 [*] |
| 4. MAI | — | — | — | — | .26 [†] | -.32 [†] | -.29 [†] | -.19 [*] |
| 5. NPI | — | — | — | — | — | -.20 [*] | .42 [†] | -.15 |
| 6. SC | — | — | — | — | — | — | .14 | .18 |
| 7. SL | — | — | — | — | — | — | — | .02 |
| 8. Age | — | — | — | — | — | — | — | — |

* $p < .05$, † $p < .01$, ‡ $p < .001$

Note. VS = Vengeance Scale; TTF = Tendency to Forgive Scale; ICI = Internal Control Index; MAI = Multidimensional Anger Inventory; NPI = Narcissistic Personality Inventory; SC = Self-Control Scale; SL = Revised Self-Liking Scale.

Discussion

I hypothesized that a linear combination of unforgiveness, externality of personal control, trait anger, narcissism, low self-control, high self-esteem, age, and sex could predict trait vengefulness. Of these traits, only age and internality did not significantly contribute to the prediction of vengefulness, the former despite a significant positive correlation; the remaining traits accounted for more than half of the variance in vengefulness scores. The results provide support for the notion that vengefulness is not merely a response to situational stimuli but is itself an individual personality trait (Brown, 2004; Stuckless & Goranson, 1992). The results furthermore suggest, consistent with the literature, that vengefulness tends to be one component of a generally violent and aggressive personality type that is also typically characterized by unforgiveness, anger, narcissism, elevated self-esteem, and poor self-control.

Of the eight variables expected to contribute to vengefulness, trait anger accounted for the largest proportion of the variance. It is unclear whether this relation is causal or cyclical: Although angry individuals may be less likely to forgive and more likely to seek vengeance, it is just as likely that vengeance-seeking reinforces their anger. Nonetheless, anger seems to be an inseparable prerequisite for vengefulness, which is characterized by angry feelings such as bitterness, resentment, and rumination on the transgression; this

study adds to the body of literature linking the two (Berry et al., 2005; McCullough et al., 2001; Worthington & Wade, 1999). The significant main effect of self-control also suggests that the predictive power of trait anger does not merely represent a failure to regulate the kind of angry motions that lead to vengefulness; rather, it provides the motivational impetus, along with unforgiveness and narcissism, to seek vengeance.

Unforgiveness also exerted a significant main effect on the prediction of vengefulness, a relation that is certainly intuitive. Individuals seek vengeance only to the extent that they have failed to forgive their transgressors. However, as Brown (2004) noted, unforgiveness alone is not a suitable basis for predicting vengefulness. Although unforgiveness may lead to vengeance-seeking, it may also result in mere grudge-holding or avoidance of the offender. This distinction is crucial because much of the prior literature on vengefulness tended to conflate the construct with simple unforgiveness (see Brown, 2003, for a discussion of the issue). The current study confirms that other personality variables, in addition to unforgiveness, can predict vengeance-seeking attitudes or behaviors. Future research should strive to examine the interaction of unforgiveness with other variables. Unfortunately, the low statistical power associated with the significance testing of interactive terms and the ineffectiveness of mean-centering in ameliorating such power issues (Echambadi & Hess, 2007) will make such studies difficult, as they will require relatively large sample sizes

TABLE 3

Revised Regression Analysis Summary

| Variable | B | SEB | β | <i>t</i> (103) | <i>sr</i> |
|-----------|-------|------|---------|--------------------|-----------|
| Intercept | 56.86 | 1.81 | — | 31.38 [‡] | — |
| TTF | -3.77 | 1.25 | -0.25 | -3.02 [†] | -.20 |
| MAI | 0.36 | 0.09 | 0.37 | 4.04 [‡] | .26 |
| NPI | 0.69 | 0.26 | 0.22 | 2.66 [†] | .17 |
| SC | -0.24 | 0.08 | -0.22 | -3.13 [†] | -.20 |
| SL | 0.48 | 0.21 | 0.19 | 2.27* | .15 |
| Sex | 5.78 | 2.69 | 0.15 | 2.15* | — |

* $p < .05$, † $p < .01$, ‡ $p < .001$

Note. TTF = Tendency to Forgive Scale; MAI = Multidimensional Anger Inventory; NPI = Narcissistic Personality Inventory; SC = Self-Control Scale; SL = Revised Self-Liking Scale.

in order to produce meaningful results.

I also hypothesized that externality would contribute to vengefulness because previous research suggested that externals were typically predisposed to long-term psychological distress and, in some cases, criminal behavior (Bayse et al., 1992). Because vengeance-seeking is one possible method of reducing the distress that results from suffering an interpersonal transgression (Worthington & Wade, 1999), there was good reason to suspect that externals would be disproportionately likely to pursue it. However, the externality term had virtually no effect on the prediction of vengefulness, nor was it significantly correlated with vengefulness. Given the measure's adequate internal consistency and the low significance values observed, it is unlikely that either measurement error or inadequate power alone can account for this finding. The likeliest explanation is that internality is as conducive to vengefulness as externality. Internal individuals may well be empowered by their sense of self-efficacy and personal control to seek vengeance, similar to the manner in which externals' desire to reduce their feelings of helplessness and lack of control might produce vengeance-seeking behaviors. Klandermans (1983) suggested just such a distinction in an effort to explain political activism in terms of locus of control theory. He proposed that internals act to cement their feelings of personal control, whereas externals act to promote feelings of personal control. Crucially, he noted that behavioral differences should emerge only when particular individuals value the behavioral outcome and the resulting feelings of control. These hypotheses, if they apply to other forms of behavior than political action-taking, suggest that locus of control is not sufficient to predict vengefulness because internal or external individuals must already value vengeance in order to pursue it as a means of satisfying their desire for personal control. Bayse et al.'s (1992) finding that incarcerated criminals exhibited higher levels of externality may, then, simply be a product of the captive environment to which they are subject.

Consistent with Brown's (2004) findings, narcissism contributed significantly to the prediction of vengefulness. Brown posited that narcissism enables vengefulness because narcissists' inflated sense of self-worth leads them to believe they are capable of achieving revenge. Exline et al. (2004) likewise suggested that narcissistic self-entitlement leads individuals high in narcissism to demand retribution after suffering a transgression. As with trait anger and unforgiveness, narcissism appears to provide one motivational impetus to enacting vengeance by removing some of the social, affective, and cognitive barriers to its pursuit. Self-esteem, independent of narcissism, also

contributed significantly to the prediction of vengefulness. Although elevated self-esteem is characteristic of narcissism, it must be noted that narcissistic self-esteem tends to be brittle and prone to fluctuation, especially in the presence of ego threats (Bushman & Baumeister, 1998; Exline et al., 2004; Rhodewalt, Madrian, & Cheney, 1998), whereas self-esteem is a more globally stable construct. Thus, even in the absence of narcissistic extremes, elevated self-esteem can promote vengefulness, presumably because individuals with low self-esteem are less likely to feel themselves capable of enacting vengeance.

The significant influence of self-control on the prediction of vengefulness is consistent with studies that found that participants whose self-control had been experimentally depleted tended to react more aggressively to insulting provocations than those whose self-control had not been so depleted (Stucke & Baumeister, 2006). This study, however, dealt with situations in which state self-control was low; the current study found that low *trait* self-control similarly related to vengefulness, a relatively novel finding. The negligible relation between self-control and forgiveness (similar to that observed by Tangney et al., 2004) suggests that highly self-controlled individuals are as likely to forgive as not. In the case of unforgiving individuals, then, self-control appears to be one of the major barriers to vengefulness: When both forgiveness and self-control are low, vengeance-seeking attitudes and behaviors become more likely, especially where anger or narcissism provides the motivation.

As noted previously, age correlated negatively with vengefulness, consistent with prior findings (e.g., Cota-McKinley et al., 2001). However, once I controlled for the other variables in the model, age contributed virtually nothing to the prediction of vengefulness. The significant negative correlation between age and trait anger suggested, and the significant Goodman test statistic confirmed, that the relation between age and vengefulness can be attributed to the decreased tendency to experience anger with age. This finding is consistent with that of Blanchard-Fields and Coats (2008), who observed that older adults tended to experience less anger than younger adults. Schieman (1999) attributed this difference to younger adults' lower satisfaction with their social and financial arrangements compared to older adults. It is unclear to what extent cultural or generational issues may play a role in these groups' respective assessments of life satisfaction; future research should endeavor to answer such questions.

Sex also significantly predicted vengefulness, with men more vengeful than women, consistent with prior studies on the topic (e.g., Brown, 2004; Cota-McKinley et al., 2001). Knox, Breed, and Zusman (2007) sug-

gested that social learning theory accounts for this difference, as men are socialized to respond with angry, vengeful aggression when emotionally injured, whereas women are conditioned to respond with sorrow. Baumeister (2007) proposed that elevated male aggression has its roots in evolutionary psychology: More aggressive men have historically been more likely to attain more sexual partners and therefore to produce more offspring. Other research has linked male aggression to the presence of elevated levels of testosterone (Carré, Putnam, & McCormick, 2009). Whether biological, evolutionary, or cultural in origin, however, the consensus in the literature seems to favor the notion that men are predisposed to being more vengeful and aggressive than women, and the current study supported that idea.

This study established that personality traits account for a sizable proportion of the variance in vengefulness scores. Thus, there is considerable room to research what other traits might similarly contribute to vengefulness. Such traits might include emotional intelligence, fearfulness, interpersonal rejection sensitivity, introversion, and guilt/shame proneness, as each governs the way individuals conduct, perceive, and process social interactions (Worthington & Wade, 1999). It does not appear that any of these traits have previously been investigated.

The significant sex term also suggests the influence of nonpersonality demographic traits on vengefulness. For example, Cota-McKinley et al. (2001) found that religious background significantly predicted vengefulness, with Biblical literalists more likely to endorse vengeance than nonliteralists. Race may also play a role, with Pettit (1987) finding that racial differences partly accounted for the probability that individuals will use arson as a means of revenge. Leach and Spears (2008) explained *schadenfreude*, the tendency to delight in the misfortune of others, as a function of a sense of personal inferiority, potentially indicating that conditions such as low socioeconomic status and other forms of out-group identification may contribute to vengefulness.

As with most correlational research, caution should be exercised in interpreting these results in terms of causality. Brown (2004) recommended that future research in the area of vengefulness attempt the replication of these results in a more controlled setting, especially one in which researchers systematically manipulate transgressions and examine resultant levels of vengefulness. That recommendation is echoed here. In addition, although the current research focused exclusively on global dispositions and personality traits, there is a relative paucity of research on the situational contexts that encourage vengeful behaviors.

Specifically, although several researchers (e.g., Brown & Phillips, 2005) have hypothesized that relationship closeness and commitment, offense severity, and other situational variables might mitigate vengeful behaviors, no empirical verification for these hypotheses has been forthcoming. As Brown and Phillips (2005) pointed out, however, examinations of these hypotheses would require the manipulation of offenses that might need to reach ethically questionable extremes in order to generate useful data. Nonetheless, such research would fill a gap in the literature that has heretofore remained empty.

References

- Aiken, L. S., & West, S. G. (1991). *Multiple regression: Testing and interpreting interactions*. Newbury Park, CA: Sage.
- Al-Mabuk, R. H., Enright, R. D., & Cardis, P. A. (1995). Forgiveness education with parentally love-deprived late adolescents. *Journal of Moral Education, 24*, 427-444.
- Baumeister, R. F. (2007, August). *Is there anything good about men?* Presented at the American Psychology Association, San Francisco, CA.
- Baumeister, R. F., Vohs, K. D., & Tice, D. M. (2007). The strength model of self-control. *Current Directions in Psychological Science, 16*, 351-355.
- Bayse, D. J., Allgood, S. M., & Van Wyk, P. C. (1992). Locus of control, narcissism, and family life education in correctional rehabilitation. *Journal of Offender Rehabilitation, 17*, 47-64.
- Belsley, D. A. (1984). Demeaning conditioning diagnostics through centering. *The American Statistician, 38*, 73-77.
- Benassi, V. A., Sweeney, P. D., & Dufour, C. L. (1988). Is there a relation between locus of control orientation and depression? *Journal of Abnormal Psychology, 97*, 357-367.
- Berry, J. W., Worthington, E. L., Jr., O'Connor, L. E., Parrott, L., III, & Wade, N. G. (2005). Forgiveness, vengeful rumination, and affective traits. *Journal of Personality, 73*, 183-225.
- Berry, W. D., & Feldman, S. (1985). *Multiple regression in practice*. London: Sage.
- Blanchard-Fields, F., & Coats, A. H. (2008). The experience of anger and sadness in everyday problems impacts age differences in emotion regulation. *Developmental Psychology, 44*, 1547-1556.
- Brown, R. P. (2003). Measuring individual differences in the tendency to forgive: Construct validity and links with depression. *Personality and Social Psychology Bulletin, 29*, 759-771.
- Brown, R. P. (2004). Vengeance is mine: Narcissism, vengeance, and the tendency to forgive. *Journal of Research in Personality, 38*, 576-584.
- Brown, R. P., & Phillips, A. (2005). Letting bygones be bygones: Further evidence for the validity of the Tendency to Forgive scale. *Personality and Individual Differences, 38*, 627-638.
- Brown, R. P., & Zeigler-Hill, V. (2004). Narcissism and the non-equivalence of self-esteem measures: A matter of dominance? *Journal of Research in Personality, 38*, 585-592.
- Bushman, B. J., & Baumeister, R. F. (1998). Threatened egotism, narcissism, self-esteem, and direct and displaced aggression: Does self-love or self-hate lead to violence? *Journal of Personality and Social Psychology, 75*, 219-229.
- Carré, J. M., Putnam, S. K., & McCormick, C. M. (2009). Testosterone responses to competition predict future aggressive behaviour at a cost to reward in men. *Psychoneuroendocrinology, 34*, 561-570.
- Cota-McKinley, A. L., Woody, W. D., & Bell, P. A. (2001). Vengeance: Effects of gender, age, and religious background. *Aggressive Behavior, 27*, 343-350.
- Duckworth, A. L., & Seligman, M. E. P. (2006). Self-discipline gives girls the edge: Gender in self-discipline, grades, and achievement test scores. *Journal of Educational Psychology, 98*, 198-208.
- Duttweiler, P. C. (1984). The Internal Control Index: A newly

- developed measure of locus of control. *Educational and Psychological Measurement*, 44, 209-221.
- Echambadi, R., & Hess, J. D. (2007). Mean-centering does not alleviate collinearity problems in moderated multiple regression models. *Marketing Science*, 26, 438-445.
- Emmons, R. A. (1987). Narcissism: Theory and measurement. *Journal of Personality and Social Psychology*, 52, 11-17.
- Enright, R. D., & The Human Development Study Group. (1991). The moral development of forgiveness. In W. M. Kurtines & J. L. Gewirtz (Eds.), *Handbook of moral behavior and development* (Vol. 1, pp. 123-152). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Exline, J. J., Baumeister, R. F., Bushman, B. J., Campbell, W. K., & Finkel, E. J. (2004). Too proud to let go: Narcissistic entitlement as a barrier to forgiveness. *Journal of Personality and Social Psychology*, 87, 894-912.
- Gunther, K. C., Cohen, L. H., & Armeli, S. (1999). The role of neuroticism in daily stress and coping. *Journal of Personality and Social Psychology*, 77, 1087-1100.
- Holbrook, M. I., White, M. H., & Hutt, M. J. (1995). The Vengeance Scale: Comparison of groups and an assessment of external validity. *Psychological Reports*, 77, 224-226.
- Hutt, M. J., Iverson, H. L., Bass, H., & Gayton, W. F. (1997). Further validation of the Vengeance Scale. *Psychological Reports*, 80, 744-746.
- Klendermans, P. G. (1983). Rotter's I. E. Scale and socio-political action-taking: The balance of 20 years of research. *European Journal of Social Psychology*, 13, 399-415.
- Knox, D., Breed, R., & Zusman, M. (2007). College men and jealousy. *College Student Journal*, 41, 494-498.
- Kroner, D. G., Reddon, J. R., & Serin, R. C. (1992). The Multidimensional Anger Inventory: Reliability and factor structure in an inmate sample. *Educational and Psychological Measurement*, 52, 687-693.
- Leach, C. W., & Spears, R. (2008). "A vengefulness of the impotent": The pain of in-group inferiority and schadenfreude toward successful out-groups. *Journal of Personality and Social Psychology*, 95, 1383-1396.
- Lefcourt, H. M. (1976). *Locus of control: Current trends in theory and research*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- McCullough, M. E., Bellah, C. G., Kilpatrick, S. D., & Johnson, J. L. (2001). Vengefulness: Relationships with forgiveness, rumination, well-being, and the Big Five. *Personality and Social Psychology Bulletin*, 27, 601-610.
- McCullough, M. E., Emmons, R. A., Kilpatrick, S. D., & Mooney, C. N. (2003). Narcissists as "victims": The role of narcissism in the perception of transgressions. *Personality and Social Psychology Bulletin*, 29, 885-893.
- McCullough, M. E., Worthington, E. J., Jr., & Rachal, K. C. (1997). Interpersonal forgiving in close relationships. *Journal of Personality and Social Psychology*, 73, 321-336.
- Meyers, L. S., & Wong, D. T. (1988). Validation of a new test of locus of control: The Internal Control Index. *Educational and Psychological Measurement*, 48, 753-761.
- Muraven, M. (2008). Autonomous self-control is less depleting. *Journal of Research in Personality*, 42, 763-770.
- Muraven, M., Pogarsky, G., & Shmueli, D. (2006). Self-control depletion and the general theory of crime. *Journal of Quantitative Criminology*, 22, 263-277.
- O'Sullivan, G. H., & Kelleher, M. J. (1987). A study of firesetters in the south-west of Ireland. *British Journal of Psychiatry*, 151, 818-823.
- Pettway, L. E. (1987). Arson for revenge: The role of environmental situation, age, sex, and race. *Journal of Quantitative Criminology*, 3, 169-184.
- Porporino, F. J., Doherty, P. D., & Sawatsky, T. (1987). Characteristics of homicide victims and victimizations in prisons: A Canadian historical perspective. *International Journal of Offender Therapy and Comparative Criminology*, 31, 125-136.
- Raskin, R. N., & Hall, C. S. (1979). A Narcissistic Personality Inventory. *Psychological Reports*, 45, 590.
- Raskin, R., & Hall, C. S. (1981). The Narcissistic Personality Inventory: Alternate form reliability and further evidence of construct validity. *Journal of Personality Assessment*, 45, 159-162.
- Rhodewalt, F., Madrian, J. C., & Cheney, S. (1998). Narcissism, self-knowledge organization, and emotional reactivity: The effect of daily experiences on self-esteem and affect. *Personality and Social Psychology Bulletin*, 24, 75-87.
- Rogge, R. D., & Bradbury, T. N. (1999). Till violence does us part: The differing roles of communication and aggression in predicting adverse marital outcomes. *Journal of Consulting and Clinical Psychology*, 67, 340-351.
- Schieman, S. (1999). Age and anger. *Journal of Health and Social Behavior*, 40, 273-289.
- Scully, D., & Marolla, J. (1985). "Riding the bull at Gilley's": Convicted rapists describe the rewards of rape. *Social Problems*, 32, 251-263.
- Siegel, J. M. (1986). The Multidimensional Anger Inventory. *Journal of Personality and Social Psychology*, 51, 191-200.
- Stucke, T. S., & Baumeister, R. F. (2006). Ego depletion and aggressive behavior: Is the inhibition of aggression a limited resource? *European Journal of Social Psychology*, 36, 1-13.
- Stuckless, N., & Goranson, R. (1992). The Vengeance Scale: Development of a measure of attitudes toward revenge. *Journal of Social Behavior & Personality*, 7, 25-42.
- Tafarodi, R. W., & Swann, W. B., Jr. (2001). Two-dimensional self-esteem: Theory and measurement. *Personality and Individual Differences*, 31, 653-673.
- Tangney, J. P., Baumeister, R. F., & Boone, A. L. (2004). High self-control predicts good adjustment, less pathology, better grades, and interpersonal success. *Journal of Personality*, 72, 271-322.
- Tice, D. M., Baumeister, R. F., Shmueli, D., & Muraven, M. (2007). Restoring the self: Positive affect helps improve self-regulation following ego depletion. *Journal of Experimental Social Psychology*, 43, 379-384.
- White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroscedasticity. *Econometrica*, 48, 817-838.
- Wilkowski, B. M., & Robinson, M. D. (2007). Keeping one's cool: Trait anger, hostile thoughts, and the recruitment of limited capacity control. *Personality and Social Psychology Bulletin*, 33, 1201-1213.
- Wilkowski, B. M., & Robinson, M. D. (2008). Guarding against hostile thoughts: Trait anger and the recruitment of cognitive control. *Emotion*, 8, 578-583.
- Worthington, E. L., Jr., & Wade, N. G. (1999). The psychology of unforgiveness and forgiveness and implications for clinical practice. *Journal of Social and Clinical Psychology*, 18, 385-418.

A Preliminary Examination of the Effects of Behavioral Descriptions on On-Task Behavior

KRISTIN B. LEMASTER

STEPHANIE M. WAGNER

ASHLEY B. TEMPEL

CHERYL B. MCNEIL*

West Virginia University

Parent-Child Interaction Therapy (PCIT) is an empirically-supported treatment containing several skill components that have not been individually examined. Behavioral descriptions are one component hypothesized to improve children's attention-to-task. Researchers coded 3 children (ages 3-6) exhibiting attention and hyperactivity/impulsivity problems as on- or off-task while the children completed a coloring task. Researchers coded participants in conditions with or without behavioral descriptions using a single subject reversal design (ABA design). Based on visual inspection of the data, participants exhibited more on-task behavior during the behavioral descriptions component, supporting the hypothesis that behavioral descriptions may have clinical utility. Due to observed practice effects and limitations to generalizability, alternative experimental designs may provide a better method of investigating treatment components.

Parent-Child Interaction Therapy (PCIT) is an evidence-based intervention for families of young children with behavioral difficulties (Herschell, Calzada, Eyberg, & McNeil, 2002b). Based on the Hanf (1969) two-stage model, PCIT uses principles from both attachment theory and social learning theory (as cited in Eisenstadt, Eyberg, McNeil, Newcomb, & Funderburk, 1993). Attachment theory has helped mold the intervention by placing focus on a positive and secure parent-child relationship. Social learning theory contributed to the therapy by placing a focus on establishing and maintaining consistent contingencies to control the child's behavior (Herschell et al., 2002b). These theoretical approaches are reflected in two phases of treatment: Child-Directed Interaction (CDI) and Parent-Directed Interaction (PDI).

The first phase of PCIT is CDI, which focuses on enhancing the parent-child relationship and developing positive interactions through play therapy. In this phase, the therapist instructs and coaches parents to follow the lead of their child's play while using the PRIDE skills: praising their child's appropriate behavior (e.g., "I like it when you play gently with the toys"), reflecting their child's speech (e.g., repeating their child's verbalizations), imitating appropriate play, describing appropriate actions of their child (e.g., "you're picking up the red block"), and exhibiting enthusiasm during

the play. Therapists also instruct parents to refrain from asking questions, using commands, and making critical comments. CDI is complete when the parent reaches mastery of the PRIDE skills, meaning that he/she uses 10 labeled praises; 10 behavioral descriptions; 10 reflections; and fewer than 3 total questions, commands, or criticisms within a 5-min period of play therapy (Herschell, Calzada, Eyberg, & McNeil, 2002a). Although labeled praises, behavioral descriptions, and reflections are measured quantitatively during this observation, the therapist also qualitatively assesses parent use of imitation and enthusiasm. CDI also emphasizes differential attention by teaching parents to provide labeled praises of appropriate behaviors and ignore mildly inappropriate behaviors. The goals of this phase include improving the parent-child relationship, child's self esteem, and sustained attention to tasks as well as decreasing the child's anger and frustration (Hembree-Kigin & McNeil, 1995). After parents master these skills, they move on to the second phase of treatment, PDI.

In contrast to CDI, in which parents follow the child's lead, the focus of PDI is to train parents in providing effective commands and discipline. Therefore,

Author note. Address correspondence concerning this article to Kristin B. Lemaster; e-mail: kristin.lemaster@my.liu.ed

*Faculty mentor

this stage begins with giving the parents the lead in playing with the child. Therapists instruct parents to use clear, direct commands while playing with the child. After a parent gives a command, he/she must praise the child for compliance or use consistent consequences for noncompliance. The therapist coaches parents in this phase of the therapy as well, giving direction and feedback for their speech and actions. There are also specific criteria for mastering this phase of treatment (Herschell et al., 2002a). The goal of PDI is to decrease disruptive child behaviors while increasing desirable behaviors (Eisenstadt et al., 1993).

Although many investigations have examined PCIT's overall effects on various childhood problems, relatively little research has focused on evaluating the many complex components and specific stages. Eisenstadt et al. (1993) compared the order in which families received the CDI and PDI stage of treatment by randomly assigning half of the families to PDI first and half to CDI first. The researchers found that mothers who received PDI before CDI reported greater satisfaction and improvements in child conduct problems than mothers assigned to CDI first.

In addition to examining the relative order of each stage of treatment, several investigations have looked at specific PCIT components. For instance, research on the timeout component indicated that the 2-chair holding technique (a procedure developed as an alternative to standard timeout for children who escape the timeout chair; parents learn to hold their child who is seated at a second "back-up" chair) was effective in both decreasing timeout escape and improving overall behavior (McNeil, Clemens-Mowrer, Gurwitsch, & Funderburk, 1994). Other research from related fields individually examining the effects of praise on child behavior found that such verbalization increased child compliance (e.g., Bean & Roberts, 1981; Budd, Green, & Baer, 1976; Reimers et al., 1993). Researchers have paid less attention to other positive parenting skills, and it is currently unclear the degree to which behavior problems in children are affected by praise versus other aspects of general attention, such as descriptions and enthusiasm (Filcheck, McNeil, & Herschell, 2001). Filcheck and colleagues examined the influence of these factors on child behavior. Experimenters presented praise enthusiastically or unenthusiastically and presented descriptions of the children's actions unenthusiastically. Children receiving unenthusiastic descriptions exhibited higher compliance than those receiving enthusiastic praise (Filcheck et al., 2001). The researchers suggested that this finding may be related to children's past experiences with unenthusiastic statements being associated with punishment. However, these findings raise questions about the roles of both

individual skills and combined skills. In particular, enthusiasm is used in combination with other skills (i.e., praise, behavioral descriptions, and reflections), and child behavior may differ depending on whether skills are used in an enthusiastic or unenthusiastic manner.

Although Filcheck et al. (2001) did investigate enthusiasm with behavioral descriptions, it is unclear how behavioral descriptions (one of the PRIDE skills of PCIT) directly affect child behaviors aside from compliance (e.g., attention to task). Behavioral descriptions involve the parent describing the child's behavior (e.g., "you're putting the eyes on Mr. Potato Head"). Theoretically, behavioral descriptions teach children concepts, keep children in the lead of the play, and show the children that they have their parent's undivided attention. Another hypothesis suggests that behavioral descriptions improve children's focus on the task at hand, thereby increasing their attention to the task (Hembree-Kigin & McNeil, 1995). The aim of the current study was to test the last of these hypotheses by addressing how behavioral descriptions affect the on-task behavior of children. We hypothesized that a child's percentage of on-task behavior would increase in a behavioral description condition compared to a nonverbal attention condition. A secondary purpose of the current project was to pilot the procedure and materials for a larger study examining several individual and combined components of PCIT.

Method

Participants

Three children between age 3 and 6 who exhibited problems with attention or hyperactivity/impulsivity based on parental verbal report participated in the study. However, participants were not required to meet diagnostic criteria for Attention-Deficit/Hyperactivity Disorder (ADHD); none of the children had a previous diagnosis.

We recruited participants by distributing flyers to Head Start centers, medical clinics, daycares, and schools. We placed flyers in common areas and distributed them to appropriate parent-child dyads. Parents who expressed interest in the study and whose children met the inclusion/exclusion criteria were scheduled for an appointment.

Participant 1 was a 6-year-old African American boy from a low socioeconomic household (below \$20,000 per year) where the highest level of parental education was a bachelor's degree. Participant 1's parental report placed him in the normal range on the ADHD Index of the Conners' Parent Rating Scale (1997; $T = 58$).

Participant 2 was a 3-year-old Caucasian girl from a low socioeconomic household where the highest level of parental education was a high school diploma.

Participant 2's parental report also placed her in the normal range on the ADHD Index ($T = 53$).

Participant 3 was a 5-year-old Caucasian girl also from a low socioeconomic household where the highest level of parent education was a high school diploma. Participant 3's parental report placed her in the normal range on the ADHD Index as well ($T = 53$).

According to parental reports, none of these children's attention problems were better accounted for by another diagnosis. These children had never received prior treatment for attention problems, nor were they taking any ADHD medications.

Participation took approximately 2 hr including consent procedures, parental completion of measurements, warm-up periods, experimental tasks, and breaks. For participation, the parent-child dyads received \$15 gift cards to Wal-Mart, small snacks and prizes for the child, and childcare for the duration of the experiment.

Materials

A demographic form designed for this study included questions about parent, child, and family characteristics in order to describe the sample. Specifically, the form contained questions about the child's age, grade, sex, ethnicity, family income, number of members in the household, and medication status.

The Conners' Parent Rating Scale (1997) is a 27-item scale with an Oppositional Scale, a Cognitive Problems/Inattention Scale, a Hyperactivity Scale, and a Conners' ADHD Index. The scale has both adequate validity and reliability (Conners, 1997). Parents use a 4-point Likert scale to report whether a behavior is *not true at all* (0), *just a little true* (1), *pretty much true* (2), or *very much true* (3). We used scores on the ADHD Index to quantify symptoms in the sample.

To code on-task behavior ratings, we defined on-task and off-task behavior based on the Revised Edition of the School Observational Coding System (REDSOCS), an evidence-based observational coding system for assessing disruptive classroom behavior (Jacobs et al., 2000). While the child completed a coloring task, raters coded behavior as on-task or off-task. The definition for on-task behavior included attending to the coloring task (e.g., maintaining visual gaze directed toward the paper or crayons) or making appropriate motor responses (e.g., coloring, switching crayons). Raters coded the child as on-task only after he/she remained on the task for the entire 10 s interval. Behaviors coded as off-task included any behavior not included in the definition of on-task behavior at any point during the 10 s interval (e.g., getting out of seat, staring blankly away from the task, using the materials inappropriately). One or two trained researcher(s)

coded every trial, and two researchers coded 29% of interactions to ensure inter-rater reliability. These raters trained for inter-rater reliability through videotaped coding of child behavior and live coding of a volunteer role-playing participant. Following training, inter-rater reliability was established through live coding and assessed by computing the percentage of agreement. Percent agreement values for these trials ranged from 83.3% to 100.0% ($M = 90.1\%$, $SD = 0.06\%$).

Design and Procedure

The experimental method of this study was a single-subject reversal (ABA) design. Each participant experienced both conditions. Prior to the beginning of each condition, the participant received a 3-min warm-up trial to control for any novelty effects. Condition A was the nonverbal attention condition, in which the participant performed a coloring task while the experimenter silently sat at a close proximity to the child to control for the effects of nonverbal attention on child behavior. Condition B was the behavioral description condition, in which the participant performed the same coloring task while the experimenter gave two behavioral descriptions for on-task behavior within a 10 s interval (e.g., "You are picking up the blue crayon; you are coloring a blue triangle"). This rate is considerably higher rate than standard PCIT coding situations (approximately 10 behavioral descriptions per 5 min or 2 per min). The purpose of this higher "dosage" was to create a more noticeable difference between the two conditions by providing a high frequency of behavioral descriptions in Condition B. When the participant was off-task in Condition B, the experimenter described the child's behaviors approximating on-task behavior. This procedure ensured a difference between the two conditions, as the presence of behavioral descriptions was the independent variable. For example, if a child were looking at the task but not coloring, the experimenter would say, "you're looking at the paper."

To measure on-task behavior, we evaluated the child's behavior during the completion of a coloring task. The child received four crayons (red, green, yellow, and blue) and a coloring sheet of geometric shapes with the appropriate colors marked in each shape. The child also received verbal instructions on how to complete the task and to continue working until told to stop. If the child asked questions during a trial, he/she received a verbal prompt to continue coloring until told to stop. We gave maximum of three verbal prompts per trial.

In both Conditions A and B, the child received crayons and the coloring sheet. The experimenter gave the child instructions and modeled the task before each trial. Rater(s) completed manipulation checks to ensure

the experimenter's adherence to the protocol (e.g., checked to ensure the correct number of descriptions was given). Raters completed manipulation checks for 68% of the trials; adherence to the protocol was 100% for all checks. Each rater simultaneously observed the experimenter's adherence to protocol and the child's on-task behavior. The experimenter did not rate on-task behavior. Each trial included eighteen 10-s intervals, thus totaling 3 min per trial. After 9 min (three 3-min trials), the child received a break. To switch conditions, on-task behavior had to remain stable (e.g., absence of bounce or trend in the direction of the hypothesis) for a minimum of three 3-min trials before switching conditions. Switching from one condition to another should be contingent on predetermined stability criterion (Barlow, Nock, & Hersen, 2009). To illustrate, if a child in Condition A was exhibiting a lower percentage of on-task behavior with each trial (i.e., downward trend), an upward trend following the switch to Condition B, with all other variables staying constant, suggests that the presentation of behavioral descriptions influenced the increase in on-task behavior. Total experimental time for each participant was dependent on how quickly the participant achieved the stability criteria outlined previously. For instance, when on-task behavior was stable for three 3-min trials in each of the three conditions, the minimum total experimental coding time was 27 min plus 9 min of warm-up.

Participants received conditions in a counterbalanced order to control for order effects. The first and third participants began with Condition B, switched to Condition A, and then switched back to Condition B (BAB), and the second participant received the opposite sequence (ABA).

Due to the directional nature of the hypothesis, trends in the opposite direction of the hypothesis were not examined as study findings. Specifically, a directional hypothesis that on-task behavior would increase in the behavioral descriptions condition compared to the nonverbal attention condition only allows the presence or absence of trends in the direction of the hypothesis to be examined as findings. A nondirectional hypothesis that the conditions would simply differ would allow all trends to be examined.

Data Analysis

Three sections on a graph represented each condition. Raters observed and graphed on-task behavior behind a one-way mirror after each trial to measure stability in conditions to determine when to switch conditions. At least three data points were required to be stable prior to switching conditions. The participants did not view these graphs at any point during the study. Each participant had an individual graph; we did not

average data across participants. Visual inspection is sufficient to detect effect sizes and allow valid conclusions to be drawn in single-subject research (Kazdin, 2003). Specifically, raters visually analyzed the graphs of on-task behavior, comparing percentages of on-task behavior in conditions A and B for each participant. No statistical analyses were conducted in this study, which is consistent with traditional single-subject methodology (e.g., Parsonson & Baer, 1992; Perone, 1999).

Results

As illustrated in Figure 1, the results for Participant 1 demonstrated a difference in on-task behavior between the two conditions. The data indicate a small and consistent increase in on-task behavior in Condition B trials compared to Condition A. In the presence of behavioral descriptions during the first Condition B (first condition), Participant 1's on-task behavior on each trial ranged from 15-18 intervals ($M = 16.3$) out of 18 total 10-s intervals per 3 min trial or 83.3%–100% ($M = 90.7%$). During Condition A (second condition), Participant 1 received no behavioral descriptions and exhibited on-task behavior ranging from 12-13 intervals ($M = 12.3$) or 66.7%–72.2% ($M = 68.6%$) per trial. When Condition B was presented a second time (third condition), Participant 1's on-task behavior ranged from 14-15 intervals ($M = 14.66$) or 77.8%–83.3% ($M = 81.5%$) per trial. Participant 1's on-task behavior was stable prior to switching conditions within each of the three trials.

Participant 2 also exhibited a difference in on-task behavior as a result of the conditions (see Figure 2). The data from Condition B demonstrated overall higher on-task behavior than either of the Condition A situations. In the first presentation of Condition A (first condition), Participant 2 exhibited on-task behavior ranging from 12-13 intervals ($M = 12.3$) or 66.7%–72.2% ($M = 68.6%$) per trial. During Condition B (second condition), Participant 2's on-task behavior ranged from 13-15 intervals ($M = 14$) or 72.2%–83.3% ($M = 77.8%$) per trial. When Condition A was presented a second time (third condition), Participant 2's on-task behavior ranged from 8-10 intervals ($M = 9$) or 44.4%–55.5% ($M = 50.0%$) per trial.

Based on visual inspection, stability was not attained for Participant 3's on-task behavior during the third and final condition, thus resulting in inconclusive results with regard to the study hypothesis. Stability was not required to complete the final condition of the study and was only required to change conditions (i.e., from the first to second condition and from the second to third condition). During the first presentation of Condition B (first condition), Participant 3's on-task behavior ranged from 9-12 intervals ($M = 10.3$)

or 50.0%–66.7% ($M = 57.4\%$) per trial. During Condition A (second condition), Participant 3 exhibited on-task behavior ranging from 5–12 intervals ($M = 8.3$) or 27.8%–66.7% ($M = 46.1\%$) per trial. Due to the trend in the direction of the hypothesis, the conditions changed from A to B. When Condition B was presented a second time (third condition), Participant 3's on-task behavior ranged from 7–14 intervals ($M = 11$) or 38.9%–77.8% ($M = 61.1\%$) per trial. Stability was not achieved in the third condition when the design was reversed.

Discussion

This study's primary purpose was to examine the effects of the behavioral description component of PCIT on on-task behavior in young children. Based on visual inspection of the data, two of the three participants exhibited slightly more on-task behavior during the behavioral description condition than during the proximity condition. This finding was evident regardless of order of conditions (e.g., ABA versus BAB), lending preliminary support for the hypothesis that adult use (i.e., parents, experimenters, and teachers) of behavioral descriptions increases a child's attention-to-task. Although this finding is consistent with theoretical assumptions presented in the PCIT manual (Hembree-Kigin & McNeil, 1995), additional research

is needed given that the current study consisted of only 3 participants.

Stability in the third condition was not attained for the last participant, thus limiting the conclusions regarding Participant 3's on-task behavior. During the second condition alone, Participant 3 completed a total of nine trials, which may have resulted in fatigue and diminished focus in subsequent trials. Based on our observations, it is possible that fatigue and practice effects resulting from repeated experimental trials may have contributed to this failure to attain stability, thereby limiting the findings.

Additionally, several factors involving the analogue situation in this study may limit its generalizability. For example, PCIT involves parent-child interactions, but this study used experimenter-child interactions. Thus, the children may have behaved differently than in a therapy situation with their parents because in real life the child has a long-term history with the parents (Lytton, 1980). This study also provided children with two behavioral descriptions every 10 s, a considerably higher rate than standard PCIT coding situations. Therefore, descriptions in the current study may have had a larger effect on child on-task behavior because of the larger "dosage." As the effects detected in this study were small in magnitude, it is unclear whether a smaller dosage

FIGURE 1

Number of intervals Participant 1 spent on-task during proximity and behavior descriptions conditions. Vertical lines depict standard deviation of on-task intervals within each condition.

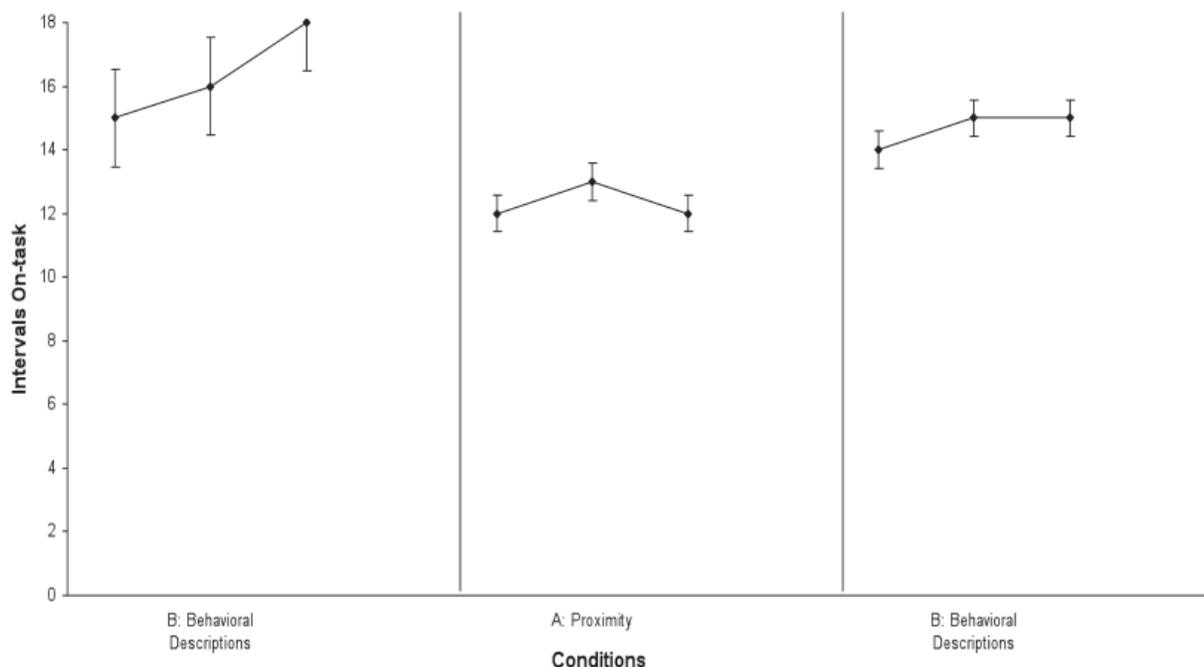


FIGURE 2

Number of intervals Participant 2 spent on-task during proximity and behavior descriptions conditions. Vertical lines depict standard deviation of on-task intervals within each condition.

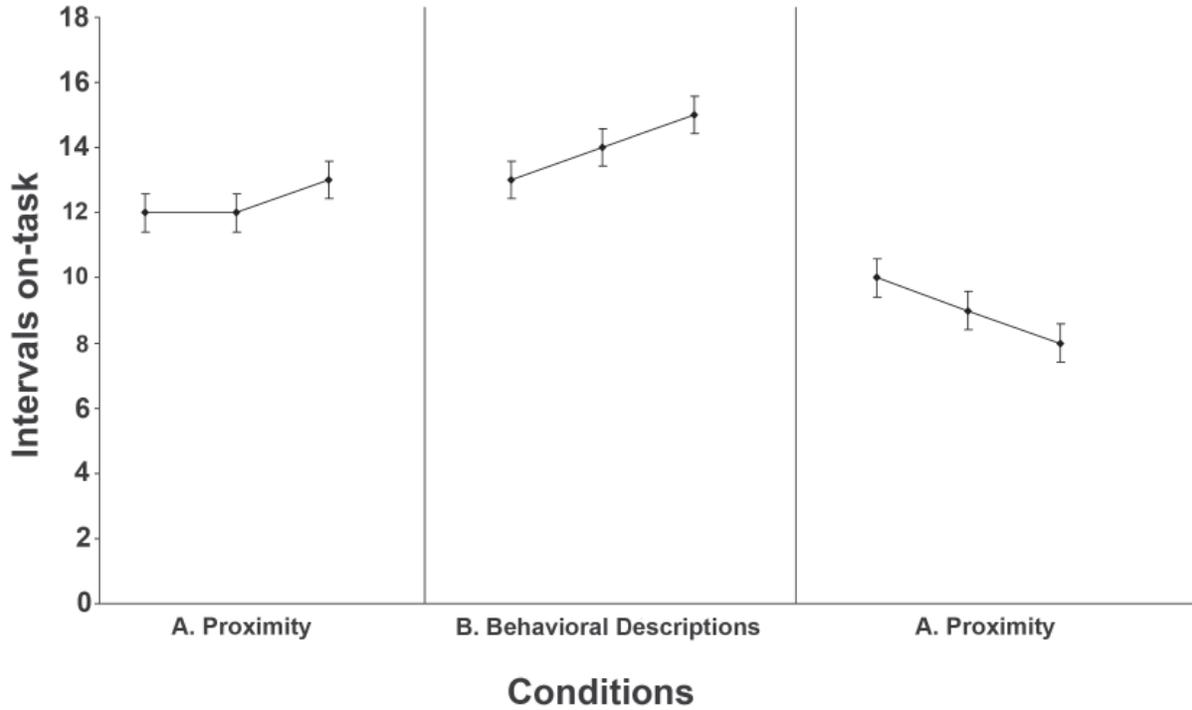
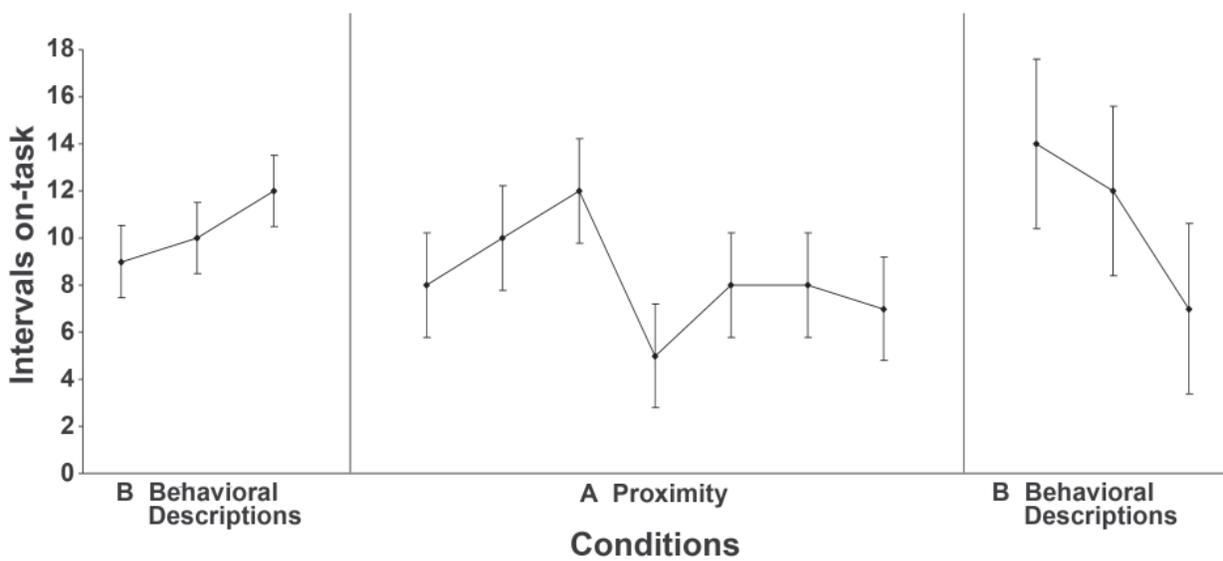


FIGURE 3

Number of intervals Participant 3 spent on-task during proximity and behavior descriptions conditions. Vertical lines depict standard deviation of on-task intervals within each condition.



would result in observable changes. Alternatively, it is also possible that the higher dosage was excessive and was less reinforcing to the child, resulting in a smaller effect than predicted. For instance, the amount of speech may have distracted the child from the task at times, and a smaller dosage may result in larger effects. Additionally, this study examined only the behavioral descriptions, whereas PCIT uses all five of the PRIDE skills simultaneously, which include three different quantifiable parental verbal behaviors (i.e., labeled praise, reflection, behavioral description). It is possible that the effects of behavioral descriptions are more powerful when used in combination with the other quantifiable parenting skills (i.e., praise, reflections). Lastly, this study recruited at-risk individuals instead of clinic-referred individuals. All three of the participants did not fall within the clinical range for ADHD on the Conners' Parent Rating Scale (1997), yet many PCIT clients meet diagnostic criteria for ADHD (Wagner & McNeil, 2008). Clinic-referred clients seeking treatment may exhibit more drastic differences in on-task behavior between conditions than the at-risk individuals examined in this study. However, the participants recruited for this experiment exhibited inattentive or hyperactive behaviors, according to parental report, which are often accompanied by oppositional tendencies (Brinkmeyer & Eyberg, 2003). Asking these active children to complete a task with specific rules and guidelines may have been measuring both child compliance and on-task behavior. Although initial compliance with the task was necessary in order to attend to the task, measuring compliance was not a purpose of the study. For instance, based on the task and coding system used in this study, a child who refused the task and a child who attempted the task but was distracted would present similarly. Therefore, it is unclear whether compliance and attention can be separated in experimental designs. These confounding variables may limit conclusions pertaining solely to the on-task behavior of these children.

Clinical Implications and Future Directions

As the results suggest, it is likely that behavioral descriptions have clinical utility, thus adding support for the theoretical assumptions underlying the PRIDE skills. This study provides preliminary evidence that behavioral descriptions help maintain on-task behavior. Further investigation is needed to better understand the scope of this skill.

Methods employed in this study are a primary area for future research. This study involved experimenter-child interactions, yet PCIT involves only interactions between parents and children. Future component analyses should explore the differences between

experimenter-child interactions and parent-child interactions and their effects on the individual components of PCIT.

Research on experimental designs for component analyses is also in need of further investigation. Based on some of the limitations of this study (e.g., fatigue and practice effects), it is possible that older children (e.g., ages 6-7) with longer attention spans could be more appropriate for this experimental design. In addition, alternative designs may provide a better method to examine components. Therefore, future studies should explore designs that reduce the number of experimental trials such as between-groups designs. However, it is also plausible that it would be difficult to detect differences using between-groups design given the small effect found in this study and the increased variability of such designs. Research should also search for a task that is more representative of demands for young children with oppositional tendencies. For example, a "free color" situation, where the child receives crayons, a blank paper, and no instructions on how to color may be more developmentally appropriate and may reduce noncompliant behavior in children.

Following further research on method and design, future research should also investigate the effects of individual PCIT components to determine which components are most crucial to the intervention. For example, continuing to separately evaluate the function of the PRIDE skills could improve therapy by emphasizing more effective components and potentially modifying less effective components. In addition to investigating individual components, the interaction between components should be explored.

Overall, this study provides some evidence supporting the clinical utility of behavioral descriptions. However, future research is needed to address current study limitations as well as investigate other components. As an intervention with multiple components, PCIT has empirical support (Brestan & Eyberg, 1998; Eyberg, Nelson, & Boggs, 2008), but further investigations of components may further improve the effectiveness and efficiency of the intervention.

References

- Barlow, D. H., Nock, M. K., & Hersen, M. (2009). *Single case experimental designs: Strategies for studying behavior change*. Boston: Pearson Education.
- Bean, A. W., & Roberts, M. W. (1981). The effect of time-out release contingencies on changes in child noncompliance. *Journal of Abnormal Child Psychology*, 9, 95-105.
- Brestan, E. V., & Eyberg, S. M. (1998). Effective psychosocial treatments of conduct-disordered children and adolescents: 29 years, 82 studies, and 5,272 kids. *Journal of Clinical Child Psychology*, 27, 180-189.
- Brinkmeyer, M. Y., & Eyberg, S. M. (2003). Parent-child interaction therapy for oppositional children. In A. E. Kazdin & J. R. Weisz (Eds.), *Evidence-based psychotherapies for children and adolescents* (pp. 204-223). New York: Guilford.

- Budd, K. S., Green, D. R., & Baer, D. M. (1976). An analysis of multiple misplaced parental social contingencies. *Journal of Applied Behavior Analysis, 9*, 459-470.
- Conners, C. K. (1997). *Conners' Rating Scales-Revised Technical Manual*. Toronto, Ontario, Canada: Multi-Heath Systems.
- Eisenstadt, T. H., Eyberg, S., McNeil, C. B., Newcomb, K., & Funderburk, B. (1993). Parent-child interaction therapy with behavior problem children: Relative effectiveness of two stages and overall treatment outcome. *Journal of Clinical Child Psychology, 22*, 42-51.
- Eyberg, S. M., Nelson, M. M., & Boggs, S. R. (2008). Evidence-based psychosocial treatments for children and adolescents with disruptive behavior. *Journal of Clinical Child & Adolescent Psychology, 37*, 215-237.
- Filcheck, H. A., McNeil, C. B., & Herschell, A. D. (2001). Types of verbal feedback that affect compliance and general behavior in disruptive and typical children. *Child Study Journal, 31*, 225-248.
- Hanf, C. A. (1969). *A two-stage program for modifying maternal controlling during mother-child (M-C) interaction*. Paper presented at the meeting of the Western Psychological Association, Vancouver, British Columbia, Canada.
- Hembree-Kigin, T. L., & McNeil, C. B. (1995). *Parent-child interaction therapy*. New York: Plenum.
- Herschell, A. D., Calzada, E. J., Eyberg, S. M., & McNeil, C. B. (2002a). Clinical issues in parent-child interaction therapy. *Cognitive and Behavioral Practice, 9*, 16-27.
- Herschell, A. D., Calzada, E. J., Eyberg, S. M., & McNeil, C. B. (2002b). Parent-child interaction therapy: New directions in research. *Cognitive and Behavioral Practice, 9*, 9-15.
- Jacobs, J. R., Boggs, S. R., Eyberg, S. M., Edwards, D., Durning, P., Querido, J. G., McNeil, C. B., & Funderburk, B. W. (2000). Psychometric properties and reference point data for the Revised Edition of the School Observation Coding System. *Behavior Therapy, 31*, 695-712.
- Kazdin, A. E. (2003). *Research design in clinical psychology* (4th ed.). Boston: Allyn and Bacon.
- Lytton, H. (1980). *Parent-child interaction: The socialization process observed in twin and singleton families*. New York: Plenum Press.
- McNeil, C. B., Clemens-Mowrer, L., Gurwitch, R. H., & Funderburk, B. W. (1994). Assessment of a new procedure to prevent timeout escape in preschoolers. *Child & Family Behavior Therapy, 16*, 27-35.
- Parsonson, B. S., & Baer, D. M. (1992). The visual analysis of data, and current research into the stimuli controlling it. In T. R. Kratochwill & J. R. Levin (Eds.), *Single-case research design and analysis: New directions for psychology and education* (pp. 15-40). Hillsdale, NJ: Erlbaum.
- Perone, M. (1999). Statistical analysis in behavior analysis: Experimental control is better. *The Behavior Analyst, 22*, 109-116.
- Reimers, T. M., Wacker, D. P., Cooper, L. J., Sasso, G. M., Berg, W. K., & Steege, M. W. (1993). Assessing the functional properties of noncompliant behaviors in an outpatient setting. *Child and Family Behavior Therapy, 12*, 93-99.
- Wagner, S.M., & McNeil, C.B. (2008). Parent-child interaction therapy for ADHD: A conceptual overview and critical review of the literature. *Child and Family Behavior Therapy, 30*, 231-256.

Self-Silencing Attitudes and Behaviors in Adolescent Couples: Links With Relationship Satisfaction

TIFFANY MICHAELSON

RENEE V. GALLIHER*

Utah State University

This study examined self-silencing attitudes and self-silencing behaviors, as they related to relationship satisfaction in 74 adolescent female/male romantic couples. We assessed self-silencing attitudes and behaviors via self-report and in the context of a problem-solving activity. Results demonstrated moderate associations between global self-silencing attitudes and reports of self-silencing behaviors in videotaped problem-solving conversations for both boyfriends and girlfriends. Further, although there were no significant associations between self-silencing attitudes and relationship satisfaction for either boyfriends or girlfriends, there were several significant moderate correlations between self-silencing behaviors and relationship satisfaction.

According to a developmental perspective, people are motivated to form close relationships with romantic partners to fulfill social and attachment needs that parents and friends previously met (Collins & Sroufe, 1999). Earlier relationships with peers and parents provided the experience by which people learned about and developed expectations about relationships. They also learned behaviors and ways of interacting that they will use in later relationships (Furman & Wehner, 1997). Researchers of romantic relationships hypothesize that romantic relationships during adolescence form the foundation for romantic experiences across the lifetime (e.g., Collins, 2003; Furman & Wehner, 1994). For this reason, understanding behaviors in healthy and maladaptive romantic relationships in adolescence may have important implications for prevention and intervention initiatives with both adolescent and adult romantic couples. In this study, we examined the psychological and behavioral construct of self-silencing as an important correlate of global relationship functioning in adolescent couples.

Self-Silencing

Researchers interested in the experience of self-silencing have identified this construct as a relationship behavior with important implications for both individual and relationship functioning. Self-silencing describes a pattern of behavior in which individuals

suppress their thoughts and opinions (Harper & Welsh, 2007) in order to preserve relationships or reduce friction within relationships. Self-silencing behaviors can be manifested with or without awareness, in that some people may intentionally choose to sacrifice their needs in order to prevent conflict or relationship distress. Others may engage in self-silencing behaviors without consciously being aware of the choice. Researchers have found links between self-silencing and depressive symptomatology (Harper & Welsh, 2007) and eating-disordered behaviors (Bucholz et al., 2007). Harper and Welsh (2007) found that adolescents with high self-silencing partners reported greater frustration and discomfort during a problem-solving interaction. They also found that girls who reported higher levels of self-silencing had lower relationship satisfaction. However, self-silencing was not related to boys' relationship satisfaction. Harper and Welsh speculated that boys may use self-silencing as a relationship management tool. As marital literature shows (Gottman, Coan, Carrere, & Swanson, 1998), men may use self-silencing or withdrawing behaviors to exercise power and control

Author note. Data for this project were collected as part of a larger study funded by a Utah State University New Faculty Grant and by R03 MH064689-01A1 from the National Institute of Mental Health, both awarded to Renee Galliher, PhD. Send correspondence concerning this article to Renee Galliher; e-mail: Renee.Galliher@usu.edu.

*Faculty mentor

in relationships, whereas women may use these behaviors to gain or maintain acceptance (Harper & Welsh). The pattern observed for girls in the Harper and Welsh study is more consistent with theoretical definitions of the self-silencing construct.

Self-silencing behaviors are particularly relevant as couples work to resolve conflicts and disagreements in their relationship. Researchers have found conflict-management behaviors, or the specific style with which a couple responds to conflict, in close relationships (Shantz, 1987) to be more relevant than the mere presence of conflict. This pattern has been shown in literature addressing both marriage (Gottman & Levenson, 2000) and adolescent relationships (Shulman, Tuval-Mashiach, Levran, & Anbar, 2006). Couples who use tactics such as compromising and negotiation to resolve disagreement and conflict tend to have longer lasting relationships (Gottman & Levenson, 2000; Shulman et al., 2006). In contrast, relationships in which individuals use conflictive or avoidant tactics do not last as long (Gottman & Levenson, 2000; Shulman et al., 2006). Researchers have also been interested in conflict tactics and how they relate to couples' relationship satisfaction. Kurdek (1995) found that couples who used withdrawing and conflict engagement had lower relationship satisfaction. Self-silencing attitudes may manifest in conflict situations as withdrawn, avoidant, or restricted engagement in the interaction. They also may be observed as yielding, giving in, or submissive behaviors. Yielding or giving in is not always negative, but according to interdependence theory (Kelly & Thibaut, 1978), relationships in which both partners are negotiating and giving in are successful. Relationships in which only one partner has a pattern of continually yielding may create power imbalances. Also, some forms of submissive behaviors have been linked with sexual exploitation and victimization (Richards, Rollerson, & Phillips, 1991).

We assessed self-silencing attitudes via self-report in a sample of adolescent romantic couples. Additionally, we assessed couple members' experiences of self-silencing behaviors in the context of a problem-solving interaction. Finally, we examined links between both attitudes and behavioral self-silencing and global relationship quality. We hypothesized that self-silencing attitudes would be manifested during the recorded conversations via more withdrawn and less connected behaviors and that both attitudinal and behavioral indices of self-silencing would be linked to lower relationship satisfaction for both boyfriends and girlfriends.

Method

Participants

Participants were 74 adolescent, heterosexual dating

couples, recruited from two rural high schools in Utah and Arizona. We randomly selected students from high school directories and contacted them by telephone to see if they were eligible (i.e., involved in a romantic relationship of at least one month duration) and interested in participating in the study. The number of adolescents initially contacted by phone was not recorded. However, the available population of students was approximately 3,000 between the two high schools. The percentage of students contacted who were both interested and eligible was very small (perhaps 10%). We sent interested adolescents information about the study and parental consent forms for both partners through U.S. mail. After one week, we contacted the target adolescent (i.e., the student originally contacted from the school directory) by phone to schedule appointments at the university research laboratory for data collection. All communication with couples was through the target adolescent, who worked with his or her partner to arrange schedules and deliver parental consent forms to the partner. Each participant provided written consent, and participants under the age of 18 provided parental consent. Participants received compensation of \$30 each (\$60 per couple).

Length of dating relationship ranged from one month to 4.5 years with the average relationship being 47.8 weeks ($SD = 37.98$) and a median relationship length of 41 weeks. The participants were between 14 and 18 years of age (girlfriends $M = 16.48$, $SD = 1.07$; boyfriends $M = 17.1$, $SD = 0.99$). Because both couple members were between 14 and 18 years of age, no couples reported age differences greater than 3 years; 84% were the same age or within one year of each other. The girlfriends were 76.4% non-Hispanic White, 18.0% Latina, 2.8% Native American, 1.4% African American, and 1.4% Asian. The boyfriends were 69.9% non-Hispanic White, 26.0% Latino/Hispanic, and 1.4% African American; one boyfriend declined to answer and one boyfriend selected "other." Because of the geographic location of data collection, approximately 70% of couple members were members of The Church of Jesus Christ of Latter-Day Saints.

Participants reported diverse family structure and educational backgrounds. Of girlfriends, 69.4% reported that their biological parents were married to each other; 80.6% of boyfriends reported that their parents were married. Fewer than 20% of adolescents reported that their parents had less than a high school degree. About 30% of both couple members' parents had completed high school. Approximately 30% of parents had a college or graduate degree, and the remainder had completed some college or technical school.

Measures

Demographic information. Participants completed a form that assessed gender, race/ethnicity, religious affiliation, educational history, parents' marital status, and parents' education.

Self-silencing. We measured attitudes about self-silencing using the 9-item Silencing the Self Subscale (STSS; Jack & Dill, 1992), which measures the extent to which participants inhibit self-expression in order to avoid conflict or possible termination of the relationship. Participants rated how strongly they agreed with each statement on a 5-point scale (1 = *strongly disagree*, 5 = *strongly agree*). Higher scores indicate higher self-silencing. Reliability analyses generated an alpha of .77 for both girlfriends and boyfriends for the self-silencing subscale.

Relationship quality. Levesque (1993) developed the Levesque Romantic Experience Questionnaire (LREQ) to measure a number of qualities in romantic relationships. We used the Relationship Satisfaction subscale to ascertain the degree to which couple members perceived their relationships as satisfying. Sample items are "In general, I am satisfied with our relationship" and "I often wish I hadn't gotten into this relationship" (reverse scored). We modified the original 6-point Likert scale to a 5-point scale (1 = *strongly disagree*, 5 = *strongly agree*) to fit the computerized administration system used in the current study. Subscale scores were calculated as the mean across the five relationship satisfaction items. Levesque found the reliability of the instrument to be high ($\alpha = .88$). The alpha calculated for the satisfaction subscale for this sample was .70 for girls and .79 for boys.

Problem-solving interaction. Couples were digitally recorded participating in three interactions drawn from previous research with adolescent romantic couples (Capaldi & Crosby, 1997; Galliher, Welsh, Rostosky, & Kawaguchi, 2004; Welsh, Galliher, Kawaguchi, & Rostosky, 1999). For the first warm-up conversation, we asked the couples to plan a party for 5 min. Then, in an effort to elicit problem-solving behaviors and conflict-management styles, for the remaining two 8-min conversations, they discussed relationship issues that they each selected from a list of common dating issues prior to video recording. Potential issues for discussion included choices such as "We don't spend enough time together" and "My parents do not like that we spend so much time together." Immediately after discussing their issues, couples completed a Global Interaction Scale adapted from Capaldi and Crosby (1997). Couple members globally rated themselves and their partners during the entire recorded interaction task on several behaviors, including "connection," "expressing true feelings," "hiding something," and "withdrawing."

Participants provided the ratings on a Likert scale (1 = *Never* and 5 = *Very Often*).

Procedures

This study used extant data from a larger study examining relationship processes in adolescent romantic relationships. Data collection took approximately 3 hr to complete, and couples participated in the university research lab of the principal investigator. We gave the couples snacks and beverages to maintain interest and attention. We digitally video recorded the couples having three conversations using a laptop computer in a room without research assistants present to provide privacy for the couples. Following recording, one couple member completed a set of questionnaire measures using a computer survey software program while the other completed a video recall procedure used for the larger study. During the final hour, couple members switched places, and each completed the remaining task. Order of survey completion (i.e., male completing video recall first or female completing video recall first) was counterbalanced to avoid order effects, and couple members completed the surveys and recall tasks in separate rooms.

Results

Preliminary and Descriptive Statistics

Table 1 displays means and standard deviations for boyfriends and girlfriends for all study variables. On average, scores for both boyfriends and girlfriends were just below the midpoint of the STSS, suggesting that couple members generally did not view themselves to be sacrificing their needs in their relationships to a great extent. There was a nonsignificant trend for a sex difference on self-silencing, with boyfriends reporting slightly higher levels than girlfriends $t(68) = -1.97, p = .052, d = .28$. Scores for relationship satisfaction were, on average, above the midpoint of the scale, and these distributions were negatively skewed. There was no sex difference in relationship satisfaction, $t(67) = .97, p = .34, d = .11$.

Means and standard deviations for boyfriends' and girlfriends' ratings of themselves for positive interaction variables (honesty, expressing true feelings, feeling connected) were, on average, above the midpoint of the scale. In contrast, scores for withdrawing and hiding something were, on average, below the midpoint of the scale. There were significant sex differences in ratings of honesty, $t(71) = 2.07, p = .04, d = .22$, with girlfriends reporting higher scores than their boyfriends. No significant sex differences were observed for feeling connected, $t(58) = .27, p = .79, d = .13$; expressing true feelings, $t(71) = 1.59, p = .12, d = .21$; withdrawing, $t(71) = .34, p = .74, d = .09$; or hiding something, $t(57)$

= -1.13, $p = .26$, $d = .11$.

Means and standard deviations for partner ratings also appear in Table 1. On average, both boyfriends and girlfriends viewed their partners as open, honest, and connected, just as they viewed themselves. Scores for withdrawing and hiding something were, on average, below the midpoint of the scale. Dependent samples t tests demonstrated no significant sex differences in any of these variables (t values ranged from -1.13 to 0.93;

p values ranged from .26 to .89; Cohen's d ranged from .00 to .18).

Associations Among Self-Silencing Attitudes and Self-Silencing Behaviors

Table 2 shows bivariate correlations examining the associations between boyfriends' and girlfriends' STSS scores and their ratings of self-silencing behaviors in the conversations. Boys who scored higher on the

TABLE 1

Descriptive Statistics

| | Boyfriends <i>M (SD)</i> | Girlfriends <i>M (SD)</i> | Possible Range |
|----------------------------|-----------------------------|------------------------------|-------------------|
| Survey Measures | | | |
| Self-Silencing | 2.77 (.60) | 2.54 (.64) | 1-5 |
| Relationship Satisfaction | 3.61 (.74) | 3.70 (.56) | 1-5 |
| Ratings of Self | | | |
| Honesty | 4.64 (.74) | 4.85 (.36) | 1-5 |
| Expressing True Feelings | 4.26 (.92) | 4.47 (.65) | 1-5 |
| Feeling Connected | 4.25 (.78) | 4.42 (.81) | 1-5 |
| Hiding Something | 1.37 (.78) | 1.30 (.60) | 1-5 |
| Withdrawing | 1.61 (.93) | 1.51 (.79) | 1-5 |
| Ratings of Partners | | | |
| Honesty | 4.66 (.59) | 4.67 (.63) | 1-5 |
| Expressing True Feelings | 4.22 (.83) | 4.37 (.93) | 1-5 |
| Feeling Connected | 4.29 (.67) | 4.32 (.84) | 1-5 |
| Hiding Something | 1.41 (.80) | 1.28 (.57) | 1-5 |
| Withdrawing | 1.49 (.84) | 1.53 (.90) | 1-5 |

STSS viewed themselves as more likely to be hiding something, with nonsignificant trends for relations with their ability to express their true feelings and withdrawal. Boys' scores on the STSS were also linked to their perceptions of the girlfriends' withdrawal during the conversation. Girls with higher scores on the STSS viewed themselves as less able to express their true feelings and less connected during their problem-solving conversations. They also rated their

boyfriends as less connected. Thus, there were some modest connections between STSS scores and ratings of self-silencing behaviors during the conversation for both couple members, particularly for self-expression and feelings of connection.

Links Between Self-Silencing and Relationship Satisfaction

Table 2 also shows bivariate correlations that assessed

TABLE 2
Bivariate Correlations of Self-Silencing, Relationship Satisfaction, and Relationship Variables

| | Boyfriend Self-Silencing | Girlfriend Self-Silencing | Boyfriend Relationship Satisfaction | Girlfriend Relationship Satisfaction |
|------------------------------|-----------------------------|------------------------------|---|--|
| Ratings of Self | | | | |
| Honesty | -.02 | -.03 | .45** | .22^ |
| Expressing True Feelings | -.22^ | -.39** | .37** | .18 |
| Feeling Connected | -.01 | -.33* | -.05 | .35** |
| Hiding Something | .29* | .21^ | -.22^ | .03 |
| Withdrawing | -.21^ | .22^ | -.18 | -.12 |
| Ratings of Partner | | | | |
| Honesty | -.06 | -.13 | .29* | .28* |
| Expressing True Feelings | .04 | -.14 | .09 | .37** |
| Feeling Connected | -.05 | -.36** | .16 | .36** |
| Hiding Something | -.03 | .12 | .04 | -.35** |
| Withdrawing | .26* | .15 | -.24* | -.17 |
| Girlfriend Self-Silencing | | | -.09 | -.12 |
| Boyfriend Self-Silencing | | | -.09 | -.12 |

^ $p < .10$, * $p < .05$, ** $p < .01$

the links between self-silencing attitudes or behaviors and relationship satisfaction. Self-silencing attitudes measured via self-report were not significantly linked to relationship satisfaction for either boyfriends or girlfriends. However, we found links between self-silencing behaviors during the conversations and self-reported relationship satisfaction. Girlfriends' perceptions of the conversation behaviors were consistently, significantly linked to their relationship satisfaction, with significant correlations observed between relationship satisfaction and girlfriends' connection, partner honesty, partner feeling connected, partner hiding something, and partner expressing true feelings. Several significant correlations emerged for boyfriends' relationship satisfaction as well, demonstrating significant relations with boyfriends' rating of their own and their partners' honesty, ratings of their girlfriends' withdrawing, and their expression of their true feelings. Thus, for both girlfriends and boyfriends, experiences of self-silencing behaviors during the conversations were more closely linked to relationship satisfaction than to their STSS scores.

Discussion

We evaluated both attitudinal and behavioral expressions of self-silencing in adolescent couples, finding that self-silencing attitudes are manifested in predicted ways when couples engage in potentially conflictual interactions. Generally, couple members in this community sample reported high levels of relationship satisfaction, which is consistent with previous research and with expectations for a healthy community sample of young couples (e.g., Galliher et al., 2004). If adolescent couples are not happy in their relationships, they simply break up. Unlike married couples, there is no social structure or cultural expectation that they maintain their relationships over time.

Couple members also reported, on average, relatively low levels of both self-silencing behaviors and attitudes in their relationships, with few significant sex differences. The nonsignificant trend for the difference between boyfriends' and girlfriends' scores on the STSS provides weak support for previous research that has observed sex differences in self-silencing attitudes (Harper & Welsh, 2007). In addition, there was a significant sex difference in partners' ratings of their own honesty, with girlfriends reporting more honesty than boyfriends. Harper and Welsh (2007) speculated that males in their study reported higher levels of self-silencing behaviors, but their reports of self-silencing may not have reflected the same kinds of submissive and self-sacrificing psychological processes proposed by Jack (1991). Although our results suggest that girlfriends' and boyfriends' levels of self-silencing attitudes

and behaviors are generally quite similar, additional research is necessary to determine whether there are differences in the *meaning* of self-silencing behaviors.

We did not find links between self-silencing attitudes and relationship satisfaction for either boyfriends or girlfriends. This finding is not consistent with past research with adolescent couples (Harper & Welsh, 2007). However, ratings of specific self-silencing behaviors were associated with relationship satisfaction. Girlfriends' satisfaction was modestly related to their experiences in the conversations, but was even more consistently and strongly related to their evaluation of their boyfriends' behaviors. In contrast, for boyfriends the strongest correlations were between their satisfaction and their ratings of their interaction behaviors. This pattern supports the notion that girls and boys are socialized to behave differently in relationships. Girls are socialized to be more relationally oriented and emotionally expressive, whereas boys are socialized to be more independent and less relationally oriented (Eagly, 1983). These findings also point to the relevance of specific interaction processes, rather than global attitudes, in developing interventions to address the potentially problematic consequences of a self-silencing relational style in adolescence.

Summary and Conclusions

There are some limitations in this study. The couple members were relatively homogenous geographically and religiously. The generalizability of the findings to couples from more diverse urban areas is questionable. Stamm (2003) argued that couples from rural areas adhere more strictly to traditional gender roles. Also, despite having a relatively substantial representation of Latino couple members in our sample, we did not have enough participants to analyze how ethnicity related to self-silencing behaviors and relationship satisfaction. Explicit examination of cultural and ethnic patterns in relationships would be an important area for future research. Finally, there is substantial debate in the literature regarding the definition and identification of adolescent romantic couples; couples in this study were required to be dating monogamously for at least one month. However, recent scholarship suggests that many important romantic experiences take place outside of the context of ongoing, committed relationships (Furman & Shaffer Hand, 2006). Allowing adolescents to define for themselves what constitutes a "romantic partner" in future research would provide a broader representation of relationships and experiences.

In summary, the results of this study support modest associations between global self-silencing attitudes and perceptions of self-silencing behaviors among adolescent romantic couple members. Further, self-

silencing behaviors, rather than global attitudes, demonstrated several significant associations with global relationship satisfaction. The developmental implications of these results for men's and women's romantic relationship functioning across the lifetime are important, as many of the patterns observed in the current study are reminiscent of findings in the adult literature. As future research yields a better understanding of self-silencing and the effects this behavior has on individuals and relationships, interventions can be designed and implemented to promote healthy functioning of adolescents and their relationships. These interventions may prevent unhealthy relationship patterns that have been found in the adult literature.

References

- Buchholz, A., Henderson, K. A., Hounsell, A., Wagner, A., Norris, M., & Spettigue, W. (2007). Self-silencing in a clinical sample of female adolescents with eating disorders. *Journal of the Canadian Academy of Child and Adolescent Psychiatry, 16*, 158-163.
- Capaldi, D. M., & Crosby, L. (1997). Observed and reported psychological and physical aggression in young, at-risk couples. *Social Development, 6*, 184-206.
- Collins, W. A. (2003). More than myth: The developmental significance of romantic relationships during adolescence. *Journal of Research on Adolescence, 13*, 1-24.
- Collins, W. A., & Sroufe, L. A. (1999). Capacity for intimate relationships: A developmental construction. In W. Furman, B. B. Brown, & C. Feiring (Eds.), *The development of romantic relationships in adolescence* (pp. 125-147). New York: Cambridge University Press.
- Eagly, A. H. (1983). Gender and social influence: A social psychological analysis. *American Psychologist, 38*, 971-981.
- Furman, W., & Shaffer Hand, L. (2006). The slippery nature of romantic relationships: Issues in definition and differentiation. In A. C. Crouter & A. Booth (Eds.), *Romance and sex in adolescence and emerging adulthood: Risks and opportunities* (pp. 171-178). Mahwah, NJ: Lawrence Erlbaum Associates.
- Furman, W., & Wehner, E. A. (1994). Romantic views: Toward a theory of adolescent romantic relationships. In R. Montemayor, G. R. Adams, & T. P. Gollota (Eds.), *Personal relationships during adolescence. Advances in adolescent development* (pp. 168-195). Thousand Oaks, CA: Sage.
- Furman, W., & Wehner, E. A. (1997). Adolescent romantic relationships: A developmental perspective. In S. Shulman & W. A. Collins (Eds.), *New directions for child development. Romantic relationships in adolescence: Developmental perspectives* (pp. 21-36). San Francisco: Jossey-Bass
- Galliher, R. V., Welsh, D. P., Rostosky, S. S., & Kawaguchi, M. C. (2004). Interaction and relationship quality in late adolescent romantic couples. *Journal of Social and Personal Relationships, 21*, 203-216.
- Gottman, J. M., Coan, J., Carrere, S., & Swanson, C. (1998). Predicting marital happiness and stability from newlywed interactions. *Journal of Marriage and the Family, 60*, 5-22.
- Gottman, J. M., & Levenson, R. W. (2000). The timing of divorce: Predicting when a couple will divorce over a 14 year period. *Journal of Marriage and the Family, 62*, 737-745.
- Harper, M. S., & Welsh, D. P. (2007). Keeping quiet: Self-silencing and its association with relational and individual functioning among adolescent romantic couples. *Journal of Social and Personal Relationships, 24*, 99-116.
- Jack, D. C. (1991). *Silencing the self: Women and depression*. Cambridge, MA: Harvard University Press.
- Jack, D. C., & Dill, D. (1992). The Silencing the Self Scale: Schemas of intimacy associated with depression in women. *Psychology of Women Quarterly, 16*, 97-106.
- Kelley, H. H., & Thibaut, J. W. (1978). *Interpersonal relations: A theory of interdependence*. New York: Wiley.
- Kurdek, L. A. (1995). Predicting change in marital satisfaction from husbands' and wives' conflict resolution styles. *Journal of Marriage and the Family, 57*, 153-164.
- Levesque, R. J. (1993). The romantic experience of adolescents in satisfying love relationships. *Journal of Youth and Adolescence, 22*, 219-251.
- Richards, L., Rollerson, B., & Phillips, J. (1991). Perceptions of submissiveness: Implications for victimization. *The Journal of Psychology: Interdisciplinary and Applied, 125*, 407-411.
- Shantz, C. U. (1987). Conflict between children. *Child Development, 58*, 283-305.
- Shulman, S., Tuval-Mashiach, R., Levran, E., & Anbar, S. (2006). Conflict resolution patterns and longevity of adolescent romantic couples: A 2-year follow-up study. *Journal of Adolescence, 29*, 575-588.
- Stamm, B. H. (2003). *Rural behavioral health care: An interdisciplinary guide*. Washington, DC: American Psychological Association.
- Welsh, D. P., Galliher, R. V., Kawaguchi, M. C., & Rostosky, S. S. (1999). Discrepancies in adolescent romantic couples' and observers' perceptions of couple interaction and their relationship to depressive symptoms. *Journal of Youth and Adolescence, 28*, 645-665.

Psi Chi Awards and Grants

| Name of Award or Grant | Submission Deadline | Who Can Apply? | Award/Grant Amount | Brief Description |
|---|-----------------------------|---|---|--|
| SuperLab Research Grants | October 1 | Graduate Undergraduate | SuperLab software Response pad | Two awards for conducting the best computer-based research. |
| Thelma Hunt Research Grants | October 1 | Faculty, Graduate, Undergraduate | Three grants \$3,000 each | Enables members to complete empirical research on a question directly related to Psi Chi. |
| Undergraduate Psychology Research Conference Grants | October 1 | Sponsor(s) of local and regional conference | Up to \$1,000 each (number varies) | Funding to defray cost of sponsoring local/regional undergraduate psychology conferences. Total grant money available is \$15,000. |
| Graduate Research Grants | November 1 February 1 | Graduate | Up to \$1,500 each (number varies) | Funding to defray the cost of conducting a research project. Total grant money available is \$20,000. |
| Undergraduate Research Grants | November 1 February 1 | Undergraduate | Up to \$1,500 each (number varies) | Funding to defray the cost of conducting a research project. Total grant money available is \$35,000. |
| Regional Research Awards | Deadlines Vary, Fall/Winter | Graduate Undergraduate | \$300 each (number varies) | Up to 78 awards presented for the best research papers submitted as Psi Chi posters for the regional conventions. |
| Denmark Faculty Advisor Award | December 1 | Faculty Advisor (chapter nomination) | Travel expense to APA + Plaque | To one outstanding faculty advisor who best achieves Psi Chi's purpose. Chapter nominates. |
| Society Annual Convention Research Awards | December 1 | Graduate Undergraduate | \$500 graduate \$300 undergraduate | Up to 16 awards (8 grad, 8 undergrad) presented for the best research papers submitted for APA/APS conventions. |
| Regional Chapter Awards | December 1 | Chapter | Twelve \$500 awards + Plaque | Presented to two chapters in each of six regions that best achieve Psi Chi's purpose. |
| Regional Faculty Advisor Awards | December 1 | Faculty Advisor (chapter nomination) | Six \$500 awards + Plaque | To six outstanding faculty advisors (one per region) who best achieve Psi Chi's purpose. |
| FBI NCAVC Internship Grants | February 1 June 1 | Graduate Undergraduate | Two grants, up to \$7,000 each | 14-week unpaid FBI NCAVC internship to conduct research; grant covers living expenses |
| Bandura Graduate Research Award | February 1 | Graduate | Travel expense to APS + Plaque + 3yr APS Membership | Student submitting best overall empirical study. Cosponsored by APS. |
| Cousins Chapter Award | February 1 | Chapter | One \$3,500 award + Travel to APA + Plaque | Presented to one chapter that best achieves Psi Chi's purpose. |
| Newman Graduate Research Award | February 1 | Graduate | Travel expense to APA + Plaque + 3yr journal subscription | Student submitting best overall empirical study. Cosponsored by APA. |
| Website Awards | February 1 | Chapter | Three \$200 awards | Presented to chapters with websites that are innovative aesthetic, and useful, and that advance Psi Chi's purpose. |
| APS Summer Research Grants | March 1 | Undergraduate | Six \$5,000 grants (\$3,500/student + \$1,500/sponsor) | Provides opportunities to conduct research during the summer with sponsors who are APS members. |
| CUR Summer Research Grants | March 1 | Undergraduate | Two \$5,000 grants (\$3,500/student + \$1,500/sponsor) | Provides opportunities to conduct research during the summer with sponsors who are CUR members. |
| SRCD Summer Research Grants | March 1 | Undergraduate | Two \$5,000 grants (\$3,500/student + \$1,500/sponsor) | Provides opportunities to conduct research during the summer with sponsors who are SDRC members. |
| Summer Research Grants | March 1 | Undergraduate | Fourteen \$5,000 grants (\$3,500/student + \$1,500/sponsor) | Provides opportunities to conduct research during the summer at recognized research institutions. |
| Kay Wilson Leadership Award | April 1 | Chapter President (chapter nomination) | One \$500 award + Travel to APA + Plaque | Award to one chapter president who demonstrates excellence in the leadership of the local chapter. |
| Allyn & Bacon Psychology Awards | May 1 | Undergraduate | 1st place—\$1,000 2nd place—\$650 3rd place—\$350 | Awards for the best overall empirical study submitted. |
| Guilford Undergraduate Research Awards | May 1 | Undergraduate | 1st place—\$1,000 2nd place—\$650 3rd place—\$350 | Awards for the overall best research papers submitted. |
| Faculty Advisor Research Grants | June 1 | Faculty Advisor | Twelve \$2,000 grants | Awards for two faculty advisors per region to conduct empirical research. |
| Model Chapter Awards | June 30 | Chapters | \$100 each chapter | All chapters meeting the five criteria will receive \$100. |

Awards and grants are submitted online at the Psi Chi website at www.psichi.org

