ABOUT PSI CHI
Psi Chi is the International Honor Society in Psychology, founded in 1929. Its mission: "recognizing and promoting excellence in the science and application of psychology." Membership is open to undergraduates, graduate students, faculty, and alumni making the study of psychology one of their major interests and who meet Psi Chi’s 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 1,150 senior colleges and universities around the world. The Psi Chi Central Office is located in Chattanooga, Tennessee. A Board of Directors, composed of psychology faculty who are Psi Chi members and who are elected by the chapters, guides the affairs of the Organization and sets policy with the approval of the chapters.

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

JOURNAL PURPOSE STATEMENT
The twofold purpose of the Psi Chi Journal of Psychological Research is to foster and reward the scholarly efforts of psychology students as well as to provide them with a valuable learning experience. The articles published in the Journal represent the work of undergraduates, graduate students, and faculty; the Journal is dedicated to increasing its scope and relevance by accepting and involving diverse people of varied racial, ethnic, gender identity, sexual orientation, religious, and social class backgrounds, among many others. To further support authors and enhance Journal visibility, articles are now available in the PsycINFO®, EBSCO®, Crossref®, and Google Scholar® databases. In 2016, the Journal also became open access (i.e., free online to all readers and authors) to broaden the dissemination of research across the psychological science community.

JOURNAL INFORMATION
The Psi Chi Journal of Psychological Research (ISSN 2325-7342) is published quarterly in one volume per year by Psi Chi, Inc., The International Honor Society in Psychology.

For more information, contact Psi Chi Central Office, Publication and Subscriptions, 651 East 4th Street, Suite 600, Chattanooga, TN 37403, (423) 756-2044. www.psichi.org; pschijsjournal@psichi.org.

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.

ADVERTISEMENTS
Advertisements that appear in Psi Chi Journal do not represent endorsement by Psi Chi of the advertiser or the product. Psi Chi neither endorses nor is responsible for the content of third-party promotions. Learn about advertising with Psi Chi at www.psichi.org/?page=Advertise

PERMISSION TO REPRINT
Permission must be obtained from Psi Chi to reprint or adapt a table or figure; to reprint quotations exceeding the limits of fair use from one source, and/or to reprint any portion of poetry, prose, or song lyrics. All persons wishing to utilize any of the above materials must write to the publisher to request nonexclusive world rights in all languages to use copyrighted material in the present article and in future print and nonprint editions. All persons wishing to utilize any of the above materials are responsible for obtaining proper permission from copyright owners and are liable for any and all licensing fees required. All persons wishing to utilize any of the above materials must include copies of all permissions and credit lines with the article submission.
184 Invited Editorial: Writing Quantitative Empirical Manuscripts With Rigor and Flair (Yes, It’s Possible)
Marianne Fallon
Central Connecticut State University

199 Self-Affirmation Intervention to Remove Negative Effects Due to Self-Objectification
Sarrah I. Ali and Heike J. M. Mahler*
University of California, San Diego

209 The Relationship Between Extraversion and Listening Comprehension Under High- and Low-Salience Visual Distraction Conditions
Nicole Virzi, Steven V. Rouse*, and Cindy Miller-Perrin*
Pepperdine University

219 Reward Responsiveness Moderates Individuals With Disordered Eating’s Implicit Attitudes Toward the Caloric Value of Food
Brittany A. Mascioli and Ron Davis*
Lakehead University

227 Doing a 180: Examining the Stability and Reversal of Behavioral Confirmation Effects
Jennifer L. Mezzapelle and Michael R. Andreychik*
Fairfield University

237 "Cna Uoy Raed Thsi Nwo? Contextual and Stimulus Effects on Decoding Scrambled Words
Sarah J. Starling* and Kelsey A. Snyder
DeSales University

251 The Effects of Perceived Attractiveness on Expected Opening Gambit Style
University of Central Arkansas
ABSTRACT. As a scientist, you are obligated to share your discoveries with colleagues and with the world. You had better do it well. In this article, I offer suggestions for writing up empirical manuscripts with quantitative data. You will learn a lot about what goes into a good manuscript, informed by the American Psychological Association’s Publication Manual (American Psychological Association, 2010) and the most recent Journal Article Reporting Standards (Appelbaum et al., 2018). Further, you will learn a good deal about how to write your manuscript so that people might enjoy reading it. Taking my cue from writing rockstars both within and beyond psychology, I encourage all scientists to adopt a classic style that puts writers and readers on a level playing field. Although I have geared this article toward emergent researchers, I hope that seasoned researchers and educators might glean something new, or at the very least, enjoy reading it.

“...you must walk up to readers and say, ‘Let’s go for a ride. You pedal, I’ll steer.” (Elbow, 1981, p. 315)

I had just completed a draft of my master’s thesis, and although I had written several APA-style manuscripts during my undergraduate days, this was the first article I had prepared for publication. I applied all the knowledge and skills I had learned and practiced. Extensive literature review? Check. Clearly presented results and conclusions? Double check. Sterling APA Style? Triple and quadruple check. I was nervous but excited to receive feedback from my graduate mentor. I waited. And waited. Finally, with a gentle smile, she returned my draft to me—steeped in red ink, blood dangling, threatening to splatter and congeal on my worn shoes.

But I had followed the rules! How could this draft go so horribly wrong? After reading my mentor’s copious feedback (which took several days), I realized that the issue was less about “what” I wrote and more about “how” I wrote it. Writing well is so much more than scholarly exhaustiveness; writing well allows you to invite readers on a journey where you serve as tour guide, passionately immersing travelers in your story without losing your companions around the bend.

Generating an empirical article—whether for a class project or for publication—feels like you are probing an alien landscape. You easily succumb to focusing on “what” because there is so much “what” to command your attention. You care less about “how” despite its ability to terraform that terrain into something habitable, maybe even beautiful. Books on academic writing (e.g., Sword, 2012) implicitly suggest that honing style is best left to the professionals with experience in such matters. But if you are to “start a stylistic revolution that will end in improved reading conditions for all (Sword, 2012, p. vii),” you had best start early. Like now, with your next manuscript—even if it is your first.

In this article, I will share insights about the what and how of quantitative manuscript writing I have learned from wrestling with peer review, reading published works on writing, and working with many undergraduates writing their first quantitative empirical articles. Full disclaimer: I could write a book on this topic; a couple of years ago, I did (Fallon, 2016). Here, I distill nuggets that I believe will help you produce a rigorous, ready-for-primetime manuscript that engages your readers.

Before diving in, know that I have made several assumptions about your research, your
experience level, and about writing manuscripts in general. First, I assume that you are writing about a single, quantitative study with a traditional IMRAD (Introduction, Method, Results, and Discussion) organization. If you have conducted multiple studies, much of this advice will still apply, but you will need to adapt accordingly. Second, I agree with Sword (2012): “academic writing is a process of making intelligent choices, not of following rigid rules” (p. 30). But you have to know the rules to make informed and reasoned choices. Your rule book is the Publication Manual of the American Psychological Association (6th Edition; American Psychological Association, 2010), which you already know because you sleep with it under your pillow. The updated Journal Article Reporting Standards (JARS; Appelbaum et al., 2018) supplement the APA Manual. Third, I expect that you are an emerging researcher, relatively early on in your journey as a psychological scientist. As such, my suggestions are not highly specialized and can be applied broadly across subdisciplines and methodologies. Fourth, I recognize that writers need to adapt their tone and style for different audiences, goals, and occasions. But good writers are card-carrying fashionistas—they are always stylish. Sacrificing accessible prose to sound scientifically rigorous is a false choice; it alienates potential future scientists (who have to start somewhere!) as well as the public. Fifth, I expect that you have come by your data honestly and ethically. No amount of sparkling prose makes up for shoddy science. And sixth, writing well takes a lot of work, a thick skin, and drive to move consistently and incrementally toward an ever-evolving, shifting, and seemingly interminable end. You will know when you have arrived—both cognitively and emotionally—at your terminus. You submit, in every sense of the word.

**Tantalizing Title**

I once likened a manuscript title to a birth announcement; you, the proud parent, get to name your baby in 12 words (Fallon, 2016). You could opt for something traditional, incorporating important constructs and teasing the finding: “Heavy Makeup Differentially Affects Men and Women’s Perception of Young Women’s Attractiveness and Confidence.” (Yes, this title is more than 12 words. Some kids have three middle names.) Traditional titling offers the advantages of sounding “scientific,” incorporates keywords that would likely cause your article to appear on database searches (Silvia, 2015), and provides enough information for readers to ascertain whether your study is relevant for their purposes. Although informative, traditional titles are not usually eye-popping.

Alternatively, you can trade some tradition for trendiness, as in “Maybe It Shouldn’t Be Maybelline: Heavy Makeup Does Not Enhance Perceptions of Young Women.” I may not know which perceptions are being studied, but I want to find out. Catchy titles do not need to follow the structure of zinger-colon-finding. Single statements with evocative language can be effective: “Heavy Makeup Can Backfire for Young Women.” Occasionally, researchers opt for questions: “Does Heavy Makeup Backfire for Young Women?”

Whimsical titles could incur costs. Cultural references may be lost on present or future readers (Silvia, 2015). Will Maybelline be around 30 years from now? Maybe not, but your manuscript will endure. Further, scholars might be turned off by cutesy titles and dismiss your work as intellectually shallow (Sword, 2012), the equivalent of people rolling their eyes when they hear you named your baby after a brand of hubcaps. Personally, I am not a fan of questions as titles. Call me old-fashioned, but I find myself answering such questions—sometimes out loud—with varying degrees of snark.

Silvia (2015) noted that your title is like a “carnival barker, luring and wheedling people inside the dark tent of your research” (p. 164). Once you have got them, you need to keep them. For that, show readers the coming attractions.

**Alluring Abstract**

An effective abstract is like a good movie trailer that draws viewers into a story and leaves them wanting more. At the same time, the abstract is a teaser with spoilers; you cannot give away the entire story, but you should deliver the highlights. You have precious little space to tell your story—most abstracts range between 150 to 250 words. So, get down to it. Quickly.

Your abstract not only provides a summary of your work, but it also persuades your readers of its importance (Sword, 2012). Use a single sentence to hook your reader. State your purpose clearly. Share compelling justifications for why your study is worth reading. Briefly describe your method (participants/sample, materials, procedure) well enough for readers to understand the basic design of your study. Highlight your most noteworthy results—findings that speak to your most central predictions or results that surprised you. Include effect sizes and confidence intervals or statistical...
Writing With Rigor and Flair

Fallon

significance levels. (Realize that including such information will inflate your word count a lot; for example, \( p = .032 \) counts as three words!) Conclude with a kicker: note practical applications of your findings, posit how your findings impact current theory, or loop back to your hook.

Given that your abstract is your manuscript’s mini-me, write it after you have drafted your manuscript. Avoid copying and pasting sentences directly from the body of your manuscript unless you want it to sound like a disjointed beat poem. That said, repeating particularly strong openers and closers could be effective. Leave your readers with an earworm that they can shake only by reading the rest of your manuscript.

Inviting Introduction

Your Introduction should bring your reader even deeper into your research, humanizing it all the while. To accomplish this, you need to harness the power of storytelling. Indeed, “To deny the power of story is to suppress our own humanity” (Sword, 2012, p. 89). Skim the introductions of the first 10 articles in a respectable journal and you will likely conclude that few psychological scientists approach manuscript writing like storytelling. Little tension, less suspense, zippo passion. You could rightly point out that these scientists nevertheless got their work published writing lifeless, turgid, yet scientifically significant prose. Why care about story? As a scientist, you are obligated to share your discoveries. As an egalitarian, you want your discoveries to be accessible. Reading an empirical article should not be an elitist rite of passage; it should be a gateway to collective understanding.

Literary storytelling involves many potent rhetorical devices (e.g., metaphor, alliteration, etc.) that normally do not wheedle their way into scientific writing. Although these devices can enliven your writing, I will focus mainly on elegant structuring and content development that will give your story a strong start. The IMRAD format imposes constraints that produce often conventional and predictable moves in empirical manuscripts (Sword, 2012). Still, mastering these conventions is not trivial. To help my students, I tell them to have a BLAST: frame the Big picture or problem; incorporate relevant Literature to contextualize your research; reveal what is Absent or lacking in said literature; briefly describe your Study; and state and rationalize Testable hypotheses. (Yes, I am this cheesy; it is part and parcel of my nerdtastic charm.)

Big Picture or Problem

Hook your reader with a juicy lede, a task harder than it looks. Sword (2012) noted that only 25% of the scholarly articles she surveyed opened with a deliberately engaging hook. The beginning of a manuscript is notorious for banal boilerplate where you can swap one construct for another. If you find yourself starting your manuscript with a variant of “Since the dawn of civilization, humankind has been fascinated by ___,” “Recently, there has been renewed interest in ___,” “Little is known about ___,” or “Merriam-Webster defines ___ as . . .” (Silvia, 2015), set fire to your computer or paper. Let the phoenix rise from the ashes.

Kail (2015) offered three solid strategies for engaging openers. You could lead with a compelling statistic to frame the problem: “Directors of counseling centers at colleges and universities report that 48.2% of their clients consider anxiety their most pressing concern (LeViness, Bershad, & Gorman, 2017).” Like a stand-up comedian, you could make an offhand observation: “Walking on a college campus is now like driving—nearly everyone is buried in their phones.” Or you could set up a compelling hypothetical situation: “Imagine you are waiting in line for coffee and someone makes a racist remark.” Starting with a quotation, as I did in this article, is another option (Sword, 2012). Silvia (2015) suggested launching with an intriguing question: “How could two people witness the same event and remember it so differently?” You have no shortage of potentially engaging openers (or duds, for that matter).

A catchy hook does not guarantee that you have reeled in your reader. Flesh out your opening paragraph with enough backstory to start humanizing your research and convincing readers why your topic is important to study (Landrum, 2008). Some writers call this initial paragraph the pre-intro or intro-to-the-intro (Silvia, 2015). Realize that your pre-intro cannot contain your entire backstory; it is the teaser for what is to come. What you choose to emphasize depends on your overarching purpose. If your study is applied, perhaps you further develop the practical ramifications for studying your topic. For studies that test basic research questions, you might tease relevant theory. Studies that fall in the middle of the applied-basic spectrum—translational research—might involve both. Does the pre-intro give away the game too early? Perhaps. You are not finding out the butler did it on the first page of a mystery novel. Rather, you should move expeditiously toward the overarching purpose of your
study, which ideally occurs at the end of your first or second paragraph. Revealing your purpose early sets the lens through which your readers frame the following paragraphs.

**Literature**

To feel connected to your characters, you want to understand their origin story. Where did they come from? What led them to their current situation? Reviewing the relevant literature is the origin story of your study. You may feel a deep personal connection with your research—and that is wonderful—but readers of empirical manuscripts need to understand the scientific backstory, rather than personal revelations of how you became interested in your research question. Hold up. Did I not just say that you need to humanize your research? Yep. But humanizing—illuminating a human connection to science, the humanity of science—is not the same as personalizing.

Incorporating relevant literature into your scientific backstory can be a tricky business. Your first concern is how much literature to incorporate, which depends on how much literature exists on your topic. Writing an exhaustive literature review can look quite scholarly, but eventually your readers will ask, “Do I really need to know all this stuff?” I am not advocating cherry-picking or choosing studies that selectively advance a singular viewpoint. I am with Kail (2015), who suggested incorporating no more than two or three citations to justify claims. A bloated glut of afterthought citations disrupts the flow of your prose (Sword, 2012) and can give the impression of smarmy namedropping. You are not trying to secure an audience with the Queen, so dial it back. (That said, I do recognize that the number of citations provides an index, albeit imperfect, of scientific impact.)

Assuming you are awash in sources, your second concern is choosing sources that best justify your claims. The type of resources matter: nearly all resources included in empirical manuscripts are scholarly, appearing in peer-reviewed journal articles, edited books, or books. Occasionally, researchers include statistics from credible web-based resources (e.g., the Center for Disease Control’s webpage) or reference the popular press, particularly when hooking the audience.

Your third, and perhaps most daunting challenge, is determining how to use your resources. Remember, your overarching goal is to provide backstory, or context for your research story. Given that you are writing an empirical manuscript, most of your backstory will be findings from empirical research. When describing such findings, it is not enough to state that a relationship has been documented. Prioritize precision in your reporting; note the direction and strength of the relationship or magnitude of the difference. Not only does this practice convey that you are a careful and thorough researcher, it opens the door to richer discussions of your results. If the current literature consistently reports a weak correlation between variables and you find a moderate to strong correlation, that is worthy of discussion. Although scholarly sources offer a trove of previous findings, they are not a one-trick pony. Scholarly sources also define theoretical constructs (e.g., conscientiousness) and describe the tenets or assumptions of a theory/model (e.g., the Big Five model of personality). An article’s methodology could inspire your own method. Or, scientists might have noted suggestions for future research or limitations that you address in your research.

Your scholarly sources are pieces of a larger puzzle; some sources provide only one piece (e.g., findings), and others offer multiple pieces (e.g., findings and theory). Fitting those pieces together into a logical and coherent picture is your fourth challenge. A tried-and-true puzzle-solving strategy is to establish the edges and then fill in. When synthesizing the background literature, your edges are major themes or claims. Let’s say you are examining whether psychological feelings of entitlement are positively related to sexism and racism in millennials (Viola & Fallon, 2018). My edges would be entitlement, sexism, racism, and the relationships among them. Consequently, I would start by describing what is known about entitlement in millennials. I would explain why entitlement should be theoretically related to sexism and racism, which means I need to define sexism and racism. Next, I would share what is known about millennials and sexism. Can you guess what would follow?

Thoughtfully using and organizing your sources reduces the likelihood that your literature review will sound like a twitchy annotated bibliography leapfrogging from one summarized article to the next. Remember, you are the filmmaker-storyteller introducing your readers to your characters. Movies that introduce a lot of characters without showing how they interact or develop will not last long in theaters, or they will go straight to DVD.

On a more local level, paraphrasing rather than directly quoting resources helps you build a coherent picture. Although it is challenging to describe
others’ research in your own words, doing so gives you much more flexibility as a writer. Think of films with characters that were inspired by characters (Romeo and Juliet) but not carbon copies of them (West Side Story). The original authors wrote with their purpose in mind and so should you. Imagine what it would be like to read a series of verbatim quotes from existing articles: all bones—no tendons, ligaments, or connective tissue.

**Absent or Lacking Knowledge**
For your research story to be compelling, you need to clearly convey your motivation for conducting your study. Even comedies need some drama to move the story forward. Expose flaws or shortcomings in the current understanding of your topic. Most of the time, authors try to hide the cracks in the vase or the stains in the carpet; here, you want to shine a spotlight on them. Doing so will convince your reader that your study is important and addresses a critical piece of the larger problem you are trying to solve. Table 1 lists some of the most common motivations for conducting empirical studies. You may find that you have more than one scientific justification for your study. Fantastic! Bludgeon your readers with this information. Otherwise, your research is nothing more than an academic exercise. Realize that you are not discussing the potential practical significance of your research here. Focus on scientific motivations, even if you are doing applied research. (Hopefully you conveyed practical significance in your hook!)

<table>
<thead>
<tr>
<th><strong>TABLE 1</strong></th>
<th><strong>Potential Scientific Justifications for Your Study</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Theme</strong></td>
<td><strong>Specific Justification</strong></td>
</tr>
<tr>
<td>Something New</td>
<td>Investigate an entirely new or understudied phenomenon</td>
</tr>
<tr>
<td></td>
<td>Validate an original questionnaire</td>
</tr>
<tr>
<td></td>
<td>Examine relationships between variables that have not been empirically linked (but are related theoretically)</td>
</tr>
<tr>
<td>Variation on a Theme</td>
<td>Test hypotheses for competing theories using a single method</td>
</tr>
<tr>
<td></td>
<td>Use different measures, stimuli, or manipulations to examine a known phenomenon or test a theory</td>
</tr>
<tr>
<td>Different Context</td>
<td>Examine known phenomena in a different population or time period (i.e., generation/era)</td>
</tr>
<tr>
<td>Wash, Rinse, Repeat</td>
<td>Provide additional evidence for a phenomenon with mixed results in the literature</td>
</tr>
<tr>
<td></td>
<td>Directly replicate a published study (especially if the findings are counterintuitive or controversial)</td>
</tr>
</tbody>
</table>

*Note. Adapted from Silva (2015) and Fallon (2016).*

**(Your) Study**
Your readers need to know the general design of your study for two reasons. First, readers need to get a sense of whether your study will fill the gap that you so skillfully exposed. If your study was designed to conceptually replicate a phenomenon, readers want to know, in broad strokes, how your study will do so. You will describe your methodology in detail in the Method; here, the extent of your description will depend on how much you think your reader needs to know. Survey research testing associations generally can be described in one or two sentences. Experimental research may need more exposition, depending on the complexity of the method. A second reason to describe your study is to help your readers better understand your hypotheses.

**Testable Hypotheses**
Your hypotheses are a contract between you and the reader—you commit to statistically evaluate each prediction you make. The most recent JARS guidelines (Applebaum et al., 2018) call for both primary and secondary hypotheses. Primary hypotheses are most central to your research question and incorporate your primary measures. Secondary hypotheses involve supplemental measures (e.g., subscales from a questionnaire) and can address potential alternative explanations for your findings such as manipulation checks for experimental designs. This practice encourages the sound scientific practice of putting your horse before your cart: you designed the study to explicitly test these predictions. In the not-too-distant past, some researchers have succumbed to polishing turds into diamonds by analyzing their findings and retrofitting their predictions, or going on statistical fishing expeditions and restructuring a study based on their most impressive catch.

State your hypotheses affirmatively: you expect a relationship between \( x \) and \( y \), that manipulating \( x \) causes a change in \( y \), that \( a \) and \( b \) uniquely explain variance in \( c \), and so on. If you are conducting null hypothesis testing, this practice runs counter to most of your training in your introductory statistics class where you were drilled to state the null hypothesis. Those of you using Bayesian approaches already state your priors affirmatively, so there is no contradiction. You can predict the direction of the relationship or effect (e.g., you expect a positive relationship between \( x \) and \( y \)) but be mindful of locking yourself into a one-tailed statistical test if you are using NHST. Hedge directionality; you can clarify your intent to use 2-tailed tests within the analysis plan of your Method section (see Field, 2018, for
reasons why you would want to use 2-tailed tests).

Word your hypotheses at the level of the variable or construct rather than at the level of the operational definition. It is much easier to understand “I expected self-esteem to be positively related to depression symptoms” than “I expected scores on the Rosenbaum self-esteem questionnaire to be positively related to scores on the Beck Depression Inventory.” For experimental enthusiasts, “I expected switch costs to increase in the presence of background noise” is more accessible than “I expected the difference in reaction time between subsequent trials in which the instructions remain the same and reaction time between subsequent trials in which the instructions differ to increase in the presence of background noise.” (Yikes.)

You do not pull your hypotheses out of thin air. Ideally, you derive hypotheses from the theory that you have summarized in your backstory. If your study is sparse on theory, your predictions should be consistent with the previous literature—and make that known. For cases where the literature has produced mixed results (i.e., some published records find the relationship, others do not or find—egads—the opposite relationship), side with the most scientifically compelling evidence.

General Organizational Advice
In my view, the Introduction is the most challenging section to write because it demands that you creatively and coherently weave ideas together, particularly for your literature review and scientific justification. The typical organizational metaphor is a funnel: You start broad and get more specific. The BLAST method provides a reasonably robust organizational template, funneling your readers from the big problem to the specific hypotheses of your study. Even so, no one-size-fits-all recipe works for all cases. Silvia (2015) suggested that your particular brand of scientific justification can affect the organizational moves you make.

To help readers keep track of where you are in your story, you have the option of including subheadings, or signposts. But beware. You could be tempted to use subheadings to subvert clear, logical transitions to the next big idea. It is like dropping stones into opposite ends of the same pond—both rocks are in the pond, but the ripples never touch.

Meticulous Method
The plot thickens as you describe your participants (or sample), materials, procedure, and data strategy in separate subsections of your Method. Although this section progresses quite linearly, you can still tell a good story (Landrum, 2008). Your priority is to provide enough detail so that readers clearly understand how you obtained your data. Your Method passes muster when: (a) your Participants or Sample subsection allows readers to make reasonable inferences about the generalizability and fidelity of your findings; (b) your Materials and Procedure subsections enable other scientists to directly replicate your study; and (c) your plan for analyzing data is sound.

Here’s the plot twist: what you include in your Method will differ dramatically depending on the nature of your data. You could be analyzing content of existing artifacts created outside a research context (i.e., magazines, songs, texts, social media posts), working with secondary data sources (i.e., analyzing an existing dataset), or collecting primary data (i.e., recruiting participants directly and contributing to a dataset). Consequently, I will address these distinct approaches—content-based datasets, secondary datasets, and primary datasets—in turn. But before I do, there are some general aspects of Method sections common to all approaches.

General Aspects of Method Sections
Traditionally, Method sections have been trisected into Participants (or Sample), Materials (or Measures or Apparatus), and Procedure. With the most recent JARS recommendations (Appelbaum et al., 2018), I suggest adding a fourth subsection: Data and Analysis Plan.

Participants. Who or what was observed, surveyed, or tested? Describe your participants (humans) or your sample (nonhuman animals or things) including relevant characteristics. Also, report exclusion criteria (i.e., how many people, nonhuman animals, or things were excluded from your sample and why they were excluded). Finally, justify your decisions about your sample size. Did you conduct a power analysis (e.g., G Power) or use another method to estimate the number of participants or observations you need to soundly examine your hypotheses? Did you achieve the sample size you intended?

Materials. What did you use to manipulate and/or measure the behavior or characteristics of the sample? When your study involves coding things or observable behaviors, you might refer to this subsection as “Coding Scheme.” At the end of the day, this subsection boils down to how you operationalized your variables. When applicable, report the construct validity of your materials (e.g.,
questionnaires, coding schemes). Also, include any specialized apparatus required to replicate your study. Typical white 8 1/2 x 11-inch paper does not qualify as specialized apparatus; finger electrodes used to measure galvanic skin response do.

**Procedure.** What specific procedures did you use to gather data? If the Materials are like ingredients in a recipe, your procedure offers chronologically ordered, specific steps to combine your ingredients. Here, your readers are looking for internal validity. And experimental design does not have a lock on it; confounding variables can insidiously slip into nonexperimental designs as well.

**Data and analysis plan.** What were your analytic plans for your data before doing any analysis? You guessed it—this section actively discourages the turds-turned-diamonds approach. Specify the conditions under which you excluded collected data: participants were not paying attention or were otherwise incapacitated (it happens!), the equipment malfunctioned, unexpected loud noises disrupted the testing session. (Notice the difference between excluding participants who did not meet inclusion criteria and excluding the data because it is not valid). If you used behavioral tasks or physiological measures that produced multiple responses per participant, describe how you intended to reduce your data and handle missing responses on trials. On tasks with multiple trials, participants occasionally goof or zone out. In such cases, you might drop those trials from analysis or replace reaction time data with an artificial maximum response time.

After describing how you would prepare your data for analysis, focus on the statistical analyses. How did you intend to statistically evaluate your primary (and secondary, if applicable) hypotheses? Which assumptions would your data need to meet (e.g., normal distribution, not heteroscedastic)? Would you transform variables to meet these assumptions? If so, how? It may seem redundant to describe what you planned to do before you analyzed your data and then report how you analyzed your data. (This is what I planned and—see?—I did it!) However, more researchers are preregistering their studies on the Open Science Framework, making their analytic plans open to increase scientific transparency. Stating how you expected to carry out your analyses provides a check against your statements on the Open Science Framework and also demonstrates that science does not always progress as planned—sometimes unexpected revelations occur as you tussle with your data.

**Specific Recommendations**

**Content-based datasets.** Begin by describing the elements that comprise your sample, how many you examined, and how you selected them. Suppose you compared 100 tweets from U.S. Presidents Barack Obama and Donald Trump. Perhaps you targeted the first consecutive 100 tweets after being inaugurated. Or perhaps you randomly selected 100 tweets from each president's first year in office. Did you consider tweet threads as a single tweet?

Next, describe how you coded these elements. Let’s say you wanted to measure optimism in those tweets. You could define which words convey optimism. You would decide whether you are counting the number of optimistic words per tweet or making a binary decision of whether the tweet sounded optimistic (or both!). Perhaps you developed a more sensitive rating scale where you subjectively rated each tweet on a scale of 0 to 10. Relay how you assessed reliability of your coding. Did you or someone else code a subset of tweets?

Within your procedure, document how you collected your data. Did you personally code each tweet, or did you use a text analysis program? If you did the coding, describe how you conducted your coding sessions: the duration of each session, how many tweets you coded in a single session, and the overall duration of coding.

**Secondary datasets.** Describe your dataset: include the name of the dataset (if it has one), when the data were collected, and how many total cases exist in the dataset. If you used only a subset of the database, what criteria did you use to select those cases? How many cases were in the subset that you analyzed?

A note about secondary data sources: Because these datasets usually have so many cases, you can conduct highly powered statistical analyses. For example, the “Emerging Adulthood Measured at Multiple Institutions 2” database (Grahe et al., 2017) is available on the Open Science Framework and contains over 3,000 cases. And it is there for the taking! However, you need to be careful. Reanalyzing the same variables in slightly different configurations inflates the likelihood of artificially finding a significant relationship between variables (a Type I error). It is like double-dipping with guacamole—the more times you go into the bowl with the same chip, the more likely you are to share your germs. The Open Science Framework makes it possible to determine which research questions others have pursued or “claimed” so you can make informed decisions. Still, you need to do your
homework and list published relevant articles using that database (see Standard 8.13 of APA’s Ethical Principles of Psychologists and Code of Conduct).

These concerns and suggestions extend to smaller nonpublic databases, including researchers who have previously amassed data and continue to draw from the same well without replenishing with new data. That well gets muddy pretty quickly. Publishing fragments from the same dataset give the erroneous impression that the data were collected afresh for each manuscript. This practice can gravely impact researchers using those sources in reviews or meta-analyses, distorting the scientific literature (APA, 2002).

In addition to documenting the characteristics of the database, you need to operationalize your variables (Materials) and describe how the database came to be (Procedure). A document containing the exact questions or measures included in the database and data collection procedures should be available for your inspection. You will not likely analyze all the variables in the database, so focus on the nondemographic variables included in your predictions. The database may include responses from multiple-item measures that need to be aggregated into a single score. For example, the EAMMi2 contains several multiple-item measures such as Brown and Ryan’s (2003) 15-item mindfulness questionnaire. Describe each multiple-item measure relevant to your study in detail and report interitem consistency (see Primary datasets below and Table 2).

Primary datasets. Here, too, start by describing your sample. State the number of people or nonhuman animals you collected data from, and include descriptive statistics of important characteristics. If you are working with humans, at a minimum note gender, age, and race/ethnic background (see Hughes, Camden, & Yangchen, 2016, for excellent suggestions regarding the collection and reporting of demographic data). Those of you working with nonhuman animals should include genus, species, and strain number (see Appelbaum et al., 2018, for other important characteristics). Summarize characteristics across groups of participants that you directly compare within your Results. For example, if your participants experienced one of three manipulated levels in an experiment, and you found substantial performance differences across groups, group-level individual differences could partially account for your findings. (This would be an example of a secondary hypothesis/analysis.) Note restrictions to your sample. If you were studying the transition to college, you would restrict your participants to first-semester students. Occasionally, (human) participants who do not match your inclusion criteria will nevertheless opt into your study. In such cases, report the number of participants you exclude for not meeting the inclusion criteria. Describe your recruitment method and sampling procedure. You do not need to trot out the jargon declaring, “I recruited a convenience sample.” Stating that you recruited human volunteers in person from an Introductory Psychology course or online through MTurk does the job. Also, if applicable, note how participants were compensated.

<table>
<thead>
<tr>
<th>TABLE 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Information to Include When Describing Different Types of Materials</td>
</tr>
<tr>
<td>Type of Material</td>
</tr>
<tr>
<td>Published questionnaires</td>
</tr>
<tr>
<td>Original or revised published questionnaires</td>
</tr>
<tr>
<td>Overt behavioral observation</td>
</tr>
<tr>
<td>Behavioral tasks</td>
</tr>
<tr>
<td>Physiological measures</td>
</tr>
</tbody>
</table>
Your materials comprise anything you intentionally expose your participants to during your study, excluding standard ethical practices such as consent forms and debriefings. Common materials include questionnaires, behavioral observations, tasks with stimuli (i.e., something—or someone—intended to provoke a response), or physiological measures (e.g., blood pressure, respiration rate). As you can imagine, these varied materials require specialized description. Table 2 lists the most important information to include when describing your materials. Because most studies have multiple materials, use subheadings to enhance organization and clarity. For example, describe each published questionnaire in a separate paragraph introduced by a subheading (Level 3 for all you APA aficionados!). Figure 1 illustrates what this structure might look like.

Start your procedure affirming that your study received approval from your institutional review board. Even if your research was exempt from IRB review, state that. Describe where you collected your data (e.g., in a laboratory; online) and how many participants were tested or observed during study sessions. For some studies, testing layout may also be important. When applicable, explain how you introduced your study to participants and obtained informed consent. Next, detail how your participants experienced your materials or—in the case of observational studies—how you observed participants. Include important details about your procedure: how you ordered a series of questionnaires, how you assigned participants to levels of a manipulated variable, how you counterbalanced within-participant conditions (if not described within your materials). You want your readers to imagine themselves as participants with under-the-hood access to the mechanics of your study. To wrap up, describe how you debriefed participants when applicable. In cases where you deceived participants, intentionally altered participants’ affective state, or exposed participants to sensitive topics, explain how you dehoaxed participants and removed or mitigated potential negative aftereffects. Finally, note the duration of your procedure.

Revealing Results

The time has come for your big reveal! Like your Method section, Results sections have a macrostructure of subsections and a microstructure of what to put in those sections. As illustrated in Figure 2, the overarching organization should be driven by your hypotheses for primary and secondary analyses and bookended by missing data and initial analyses up front, and exploratory analyses in the rear.

Missing Data

Although you described your plan for addressing missing data within your Method, now you report how many data points you actually discarded and why you discarded them. Include frequencies or percentages for each reason: “A total of 6 animals’ complete records were discarded for equipment malfunction (n = 4) and experimenter error (n = 2).” Also, state whether and how you replaced missing data. Report exactly how often you made such replacements.

Initial Analyses

Provide readers with enough information about your outcome variables to determine whether statistical assumptions have been met for your analyses. For example, you may need to examine whether continuous outcome variables are normally distributed or to ascertain how strongly predictors are correlated. Also, note outliers or points of
influence and their fate—did you exclude them from subsequent analyses? Further, did you transform the data to address nonnormality, or did you decide on a statistical test that does not require that assumption to be met (e.g., robust regression)? By looking at your data, you also might find that a continuous variable is bimodal or multimodal; no transformation will fix that, but you could consider making the variable categorical. If you did, report the ranges of scores that determined membership in each category.

Primary and Secondary Analyses
As with your Data and Analysis Plan subsection of your Method, I highly recommend using a subheading structure that reminds readers of the hypotheses you statistically evaluated (see Figure 2). For each hypothesis, begin by noting how many cases were included in the analysis. Restate the analysis you used to evaluate the prediction. Include descriptive and inferential statistics. Where appropriate, include exact $p$ values, effect sizes, and confidence intervals. Signal whether your findings are statistically significant (when applicable) and describe the direction of the relationship or effect. That’s a tall order, but you can accomplish many, if not all of these goals in a single sentence: “A 2 x 2 ANOVA using sexual debut (early, late) and biological sex (woman, man) as between-subjects variables and relationship duration as the between-participants variable revealed that women ($M = 4.36$ months, $SD = 1.93$), 95% CI [3.81, 4.92], reported longer relationships than men ($M = 3.21$ months, $SD = 1.55$), 95% CI [2.62, 3.80], $F(1, 74) = 8.54$, $p = .005$, $\eta^2 = .104$” (Vancour & Fallon, 2017, p. 127). Yes, the writing is dense. Of all the sections in your manuscript, your Results will likely be the most technical and feel the most foreign. For all my haranguing about style, here you have to play it straight.

Liberally use tables to summarize data—particularly descriptive statistics—and figures to illustrate main findings. Admit it: you have at one point skimmed (skipped?) over some text in a Results section and focused on the tables and figures. And for good reason! Vivid visuals powerfully display your findings. See Nicol and Pexman (2010a, 2010b) for multiple examples of effective tables and figures—a visual for every occasion.

Exploratory Analyses
Sometimes when you are working with a dataset, you will have a flash of insight that you did not consider when you planned the study. When this occurs, you have the option of reporting your finding as an exploratory analysis. Remember, you do not have license to fish and you should statistically correct for the additional tests you have conducted. But if you happen to find something interesting, you can report it so that you or another researcher can explicitly design a study to follow up your promising, yet preliminary results.

Discerning Discussion
Time to wrap it up and roll credits. Before that, you have some (actually, a lot of) explaining to do. Your Discussion is part extension of your Results and part inverse of your Introduction. You focus on explaining your Results, then zoom out to the big picture or problem. Your Discussion contains multiple components, most of which can be organized interchangeably—so no cheesy acronym this time. (Sorry to disappoint.) However, Discussions usually begin with a summary and end with a take-home message that speaks back to your opening hook. In the middle, you place your findings in the context of previous research, entertain alternative explanations for your findings, acknowledge limitations of your study, offer ideas for future research, and consider practical applications of your findings.

Summarize Your (Primary) Results
Your readers have just worked through multiple analyses and statistics; do them a solid and summarize the main findings of your study. I recommend focusing on primary analyses, clearly stating whether your findings were consistent with expectations. Your research story has three potential outcomes: your findings supported your hypotheses, your findings partially supported your hypotheses, or your findings did not support your hypotheses. Regardless of the endgame, thoughtfully and thoroughly discuss your findings.

Relate Findings to Previous Research
Remember that research you used to justify your predictions? Revisit these sources and connect your findings to previous research; show readers how your findings fit—or do not fit—with our current understanding. When appropriate, discuss whether your findings are consistent with the theory you used to derive predictions. Readers should come away with a clear impression of how your findings enrich collective understanding.

If you reported exploratory analyses, you likely uncovered unexpected relationships that you could
not have anticipated within your Introduction. In such cases, incorporate literature that helps readers appreciate how your preliminary findings fit into a broader context. Use the literature to inform potential explanations for your exploratory findings.

Entertain Alternative Explanations (and Acknowledge Limitations)
You may think that posing alternative explanations for your findings weakens your conclusions. But it is better to get ahead of criticism rather than letting someone else poke holes in your work. Entertaining alternative explanations demonstrates that you have thought deeply about your findings, and although you cannot address every possible alternative explanation, you should give your readers an insightful sampling.

Your secondary hypotheses and analyses point you toward alternative explanations of your findings. Indeed, you planned these analyses to rule out alternative explanations. For example, manipulation checks (when done well) demonstrate that a manipulation worked. Analyses demonstrating that the strength of the effect was correlated with the strength of the manipulation suggests that the manipulation caused the observed effect.

Acknowledging limitations—carefully considering your study’s validity from multiple angles—provides another means to address alternative explanations for your findings. No study is perfect, including yours. For example, null results happen for many reasons: no relationship actually exists; the manipulation might not have been strong enough or your measures not sensitive enough (i.e., construct validity); or uncontrolled extraneous variables wreaked havoc and obscured relationships (i.e., statistical validity). Also, discuss how generalizable your findings are outside your specific research context (i.e., external validity). Is your sample “WEIRD” (Western, educated, industrialized, rich, and democratic)? Are the demographic characteristics of your sample representative of the population from which your sample was drawn? You would not want to claim that your findings apply to all young adults when your sample is not representative of young adults.

Beware letting your negativity bias run amuck, resulting in a litany of limitations. Not all limitations are compelling. Sample size is such low-hanging fruit that my students can hardly resist stating that their sample size was not large enough despite reaching their recruitment goals for ample statistical power. Restrict your discussion of limitations to those that are truly thought-provoking and have the potential to be addressed in future research.

Propose Future Research
Here comes the fantasy sequence of your movie. Derive ideas for future research by remedying limitations, following up exploratory analyses, or taking your research to the next logical step(s). Limitations are future research ideas ripe for picking complete with stock phrase: “Future research should address this concern.” Similarly, pursuing promising preliminary findings is a gimme: “Future research should further examine these promising results.” Dreaming up next steps is the money-maker. Leverage your curiosity and engage—even surprise—your readers.

Consider Practical Implications
All research can have practical applications. But the more “basic” your research question, the more removed it is from direct application. Excluding other researchers, consider who could use your findings—Teachers? Caregivers? Health-care providers? Mental health practitioners? Also, note how these people might use your findings: “... the present results could help sex educators and clinicians counsel young people who are considering becoming sexually active within their romantic relationships” (Vancour & Fallon, 2017, p. 129).

Deliver the Take-Home Message
Every fiber of your being will want to repeat your main findings. Resist. Instead, remind readers why your findings matter. Bring them back to the big picture that you introduced in your hook. Take heed—just like your hook, it is easy to cop out with a soulless take-home message: “The present findings have important implications for ___. ” If that is your closer, fire up your preferred incendiary device. Do the hard work of developing a compelling hook and looping back to it at the conclusion of your paper. I stand by my words: “When deciding whether to give your report a thoughtful read or a cursory glance, your audience will scan the first and final paragraphs of your report. Give your readers every reason to explore all that lay between” (Fallon, 2016, p. 106).

APA Style and Format
Many researchers (including yours truly) have tried to sell the importance of writing in APA style and format (Fallon, 2016; Landrum, 2008; Silvia, 2015). Having a set format (e.g., IMRAD) may
seem constraining, but knowing the overarching structure of your manuscript takes some of the guesswork out of organizing and allows you to focus creative energies on writing stylishly within a framework.

The sheer number of formatting rules and stylistic guidelines is mindboggling. The good news is that many resources can help you. In addition to the APA Manual (APA, 2012), open resources including the APA style blog (http://blog.apastyle.org/), an editorial from this journal (Hughes, Brannan, Cannon, Camden, & Anthenien, 2017), and YouTube tutorials (Fallon, 2014a, 2014b) expose the nigglier details of APA style and format.

**How to Craft a Stylish Manuscript**

Now that you know the “whats” of a rigorous empirical manuscript, here are some “hows” to help you add flair to your research story. But first, perhaps you are wondering why style matters. Pinker (2014) offered three reasons: (a) writers who clearly transmit messages produce thankful readers who not only “get it” but are spared the migraine of wresting meaning from impenetrable prose; (b) stylish writers earn readers’ trust—writing clearly and deftly conveys that you appreciate readers’ needs and are willing to put effort into communicating; and (c) stylish writing adds beauty and joy to life.

**Classic Style**

Most scientific manuscripts are written in practical style: the writer is the expert, the reader is a noob, the writer imparts knowledge to the reader, full stop. Pinker (2014) advocated moving toward classic style (Thomas & Turner, 1994), which assumes an equal relationship between writer and reader. The writer conversationally guides the reader to see something—some truth—that the reader has already seen or, more accurately, that the reader believes the writer has seen. In a nutshell, a practical writer pours water into the reader’s glass, not asking the reader to know when to break them, and read your manuscript out loud. You will be astonished hearing what you expected your readers to just “know.”

In addition to making the abstract concrete, classic stylists are students of form and structure. As a psycholinguist, Pinker (2014) thoroughly disemboweled arcane sacred cows of grammar—split infinitives and the like—that obstruct classic style. Again, you have to think deeply about the rules to know when to break them, and this process is not intuitive: “…the unconscious mastery of language that is our birthright as humans is not enough to allow us to write good sentences” (p. 78). I do not have the luxury of discussing grammar and syntax in the writer works as hard to understand as the writer works to be understood. To infuse writing with classic style, you root content in tangible experience and establish a conversational tone.

**Content.** If the goal is to guide readers to literally see the world as the writer believes it to be, classic writers make readers visualize the
Writing With Rigor and Flair

Fallon

depth here (wait, was that a groundswell of relief?), but I implore you to deepen your relationship with language. When your relationship has fizzled, or becomes strained or downright antagonistic, Pinker (2014) is as good a couples’ counselor as they come.

**Tone.** Writing conversationally is achieved through crafting a tone that is informal, personal, collaborative, and confident (Silvia, 2015). Think of singers who give virtuosic performances—Lady Gaga, Beyoncé, P!nk. Across (and often within) songs, they mold their vocal delivery and timbre to convey contrasting emotions and evoke reactions in their audience. Similarly, effective writers deliberately consider what tone works best for a given project and flexibly adapt it to create desired effects.

You may be surprised to learn that you should be working toward a more personal tone in your manuscript. Developing a personal tone is not the same as personalizing, or disclosing intimate details of your life. But a personal tone shortens the distance between you, your work, and your readers. An impersonal tone sounds like: “When participants appeared distracted, they were sternly redirected toward the task by the experimenter.” Assuming you are the experimenter, the third-person passive construction strips you of your agency. With a more personal tone, you would get: “When participants’ attention lagged, I verbally encouraged them to refocus on the task.”

You may be further aghast to learn that your academic writing should be less formal than traditional practical style would have you expect. If you write informally, people might not take you seriously (Sword, 2012). But consider the alternative: Writing too formally turns the reader off. Who wants to invest their time reading something dry and stodgy? An informal tone makes your intellectual thought process evident and accessible. You will not hit the sweet spot by writing like you talk: “I started reading Carol Dweck’s *Mindset* like 2 weeks ago and I—ya know—and it’s been an equal—does not make you a dullard. Writing pretentiously makes you a jerk.

Writing confidently is challenging. The more you know, the more you are acutely aware of everything you do not yet know, which compels you to pepper your claims with hedging qualifiers (e.g., somewhat, slightly). You do not want to come across as knowing everything, but you do want to assertively convey that you are sharing your best ideas at the moment. You can say something confidently while qualifying your claims: “I found a statistically significant, but weak correlation between magical thinking and the number of years that participants were avowed Cubs fans.”

To assess your tone, Silvia (2015) suggested rating yourself from -10 to +10 on these four dimensions (informal, personal, collaborative, and confident) with 0 being neutral. Consistently strive for the positive side of the scale for the personal, collaborative, and confident dimensions. The dicey informal-formal dimension depends on your manuscript’s eventual outlet. Nevertheless, I would aim between +3 and +6 toward informal. Viva la revolución! If you are concerned about setting the appropriate tone, ask someone who will be honest with you to read your manuscript and rate it using these scales.

**The Eight-Item Checklist of Stylish Writing**

Sword (2012) asked 70 academics across disciplines to describe what makes academic writing stylish. She distilled these interviews into eight lessons, which make a lovely checklist:

1. Express complex ideas clearly and precisely;
2. Produce elegant, carefully crafted sentences;
3. Convey a sense of energy, intellectual commitment, and even passion;
4. Engage and hold readers’ attention;
5. Tell a compelling story;
6. Avoid jargon, except where specialized terminology is essential to the argument;
7. Provide readers with aesthetic and intellectual pleasure; and
8. Write with originality, imagination, and creative flair. (pp. 7–8)

You have already read about most if not all of these lessons within this article. But presenting these gems within this crystallized, succinct, and
Writing an empirical manuscript is a big deal, and you want to do it well. But your quest for scientific rigor should not hamstring the goal of producing stylish prose. I fear that emergent researchers find writing empirical articles yet one more way to increase their fluency in academese, writing to the 2%. Instead, you should take Dawkins (1996) literally and “try to inspire everybody with the poetry of science” (p. viii) by conducting elegant research and communicating it lyrically in classic style. Although I have not directly addressed the emotional or motivational aspects of writing, it bears mention that you do not wake up one morning suddenly capable of producing stylish and rigorous prose. You hone this skill incrementally over time and many, many drafts. You write when you do not really want to and think you are producing schlock. (Maybe you are, maybe you aren’t.) You fuss over what you write and how you write it.

With the benefit of hindsight, I can look back on receiving my thesis feedback with good humor and even nostalgia. At the time, I felt like my heart was ripped out of my chest and my brain was even nostalgia. At the time, I felt like my heart was ripped out of my chest and my brain was

**Final Thoughts**

Writing an empirical manuscript is a big deal, and you want to do it well. But your quest for scientific rigor should not hamstring the goal of producing stylish prose. I fear that emergent researchers find writing empirical articles yet one more way to increase their fluency in academese, writing to the 2%. Instead, you should take Dawkins (1996) literally and “try to inspire everybody with the poetry of science” (p. viii) by conducting elegant research and communicating it lyrically in classic style.

Although I have not directly addressed the emotional or motivational aspects of writing, it bears mention that you do not wake up one morning suddenly capable of producing stylish and rigorous prose. You hone this skill incrementally over time and many, many drafts. You write when you do not really want to and think you are producing schlock. (Maybe you are, maybe you aren’t.) You fuss over what you write and how you write it.

With the benefit of hindsight, I can look back on receiving my thesis feedback with good humor and even nostalgia. At the time, I felt like my heart was ripped out of my chest and my brain was

**Recommended Reading**


Author Note: Marianne Fallon, Department of Psychological Science, Central Connecticut State University.

I thank all the emergent researchers and writers with whom I have worked over the years. I have learned so much from you. I also thank my undergraduate mentor, Andrea Halpern, and my graduate mentor, Sandra Trehub, for helping me become a better writer.

Correspondence concerning this article should be addressed to Marianne Fallon, Department of Psychological Science, Central Connecticut State University, 1615 Stanley Street, New Britain, CT 06050. E-mail: fallonmar@ccsu.edu
It is well understood that Western culture promotes extreme thinness as the ideal female body, and this has contributed to the problem of negative body image in women (Stice, Mazotti, Weibel, & Agras, 2000). Although the body image issues of men have received less attention and are less well understood (Strother, Lemberg, Stanford, & Turberville, 2012), it is increasingly evident that men also have body image concerns (Muth & Cash, 1997). The issue of negative body image in both sexes is important because it is associated with reduced self-esteem and increased distress (Cohane & Pope, 2001). One factor that is thought to contribute to negative body image is self-objectification, which occurs when people begin to internalize an observer's perspective of their own bodies (Fredrickson & Roberts, 1997). This can be problematic because internalization leads to habitual body monitoring, wherein individuals monitor their bodies as they believe observers do, and place a greater emphasis on how they look rather than on how they feel (Fredrickson & Roberts, 1997). This in turn can lead to feelings of anxiety (Fredrickson & Roberts, 1997), body shame (Fredrickson & Roberts, 1997; Tiggemann & Williams, 2012), and is a risk factor for eating disorders (Fredrickson & Roberts, 1997; Maine & Bunnell, 2010). Further, a recent study showed that self-objectification is even positively correlated with appearance fixing (i.e., trying to change outward appearance) and avoidance coping (i.e., disengaging in potential body image threat situations), two maladaptive behaviors that have been linked with lowered self-esteem, disordered eating behaviors, and lower quality of life related to body image (Bailey, Lamarch, Gammage, & Sullivan, 2016).

A recent study by Register, Katrevich, Arguete,
Self-Affirmation to Remove Self-Objectification

Ali and Mahler

Self-Objectification negatively affects body image in both men and women. This is one of the few studies that specifically aimed to manipulate self-objectification, rather than just measure it, as well as to study both men and women. Experimenters induced a state of self-objectification in college students with a writing task asking them to describe their bodies from an observer’s viewpoint. Compared to those in the control group, those who were self-objectified scored significantly higher on a questionnaire measuring self-reported eating pathology. This suggests that those who completed the self-objectification writing task had more negative eating attitudes compared to those who were not self-objectified. Eating pathology has been shown to be a predictor of body dissatisfaction (James, Phelps, & Bross, 2001; Lawler & Nixon, 2010). Therefore, it is possible that those who completed the self-objectification writing task also felt more negatively about their bodies. To measure self-reported eating pathology, Register et al. (2015) used the Drive for Thinness subscale of the Eating Disorder Inventory (Garner, 1991).

One goal of the current study was to replicate and extend the work of Register et al. (2015) with a more appropriate assessment of body image for men. To achieve this, we added a measure of body image that specifically assesses the more muscular body preferred by some men (Cohane & Pope, 2001) and thereby should allow a more accurate understanding of the effects of self-objectification.

The primary goal of the current study was to determine whether the effects of self-objectification might be removed via self-affirmation. To our knowledge, no other study in the self-objectification literature has aimed to experimentally determine how self-objectification effects can be removed or reduced. As mentioned above, it is important to discover methods of removing the effects of self-objectification because it may lead to body shame, anxiety, and is a risk factor for eating disorders. One possible way to do this is by boosting positive feelings about the self to try to counteract the negative feelings that occur due to self-objectification. Self-affirmation may be a method of removing or reducing the negative effects of self-objectification because it has been found to effectively boost self-integrity (Steele, 1988), mood (Koole, Smeets, Van Knippenberg, & Dijksterhuis, 1999), and sometimes self-esteem (Fein & Spencer, 1997; Sherman & Cohen, 2006)—all of which could potentially be negatively affected by self-objectification.

To examine these issues, male and female college students were randomly assigned to one of six conditions in a 2 (self-objectification condition: objectified vs. not objectified) x 3 (self-affirmation condition: self-affirmation vs. control affirmation vs. no affirmation) between-subjects design. We hypothesized that women in the objectification condition who did not undergo self-affirmation would show a higher drive for thinness than those in the control condition. We also hypothesized that men in the objectification condition who did not undergo self-affirmation would report a higher drive for muscularity than those in the control condition. Further, we predicted that undergoing self-affirmation would remove the negative effects of self-objectification such that men and women in the self-affirmation condition would report a lower drive for thinness (women) and muscularity (men) than those who did not undergo self-affirmation.

Method

Participants

Participants were 125 female and 53 male undergraduates (M = 20.14 years, SD = 1.85). The ethnic/racial background of the sample was 41.6% Asian, 23% European American, 10.7% Latino-Hispanic, 1.7% African American, and 23% other or mixed ethnicity.

Measures

Body image. Body image was assessed with questions from the Drive for Thinness subscale of the Eating Disorder Inventory-2 (Garner, 1991) and the Drive for Muscularity Scale (McGreary, 2013). The Drive for Thinness subscale consists of seven items and the Drive for Muscularity Scale is composed of 15 items. An example of an item on The Drive for Thinness subscale is, “I am preoccupied with the desire to be thinner.” Responses were rated on a 1 (never) to 6 (always) Likert-type scale for both measures. Cronbach’s α for the Drive for Thinness subscale was .92, and for the Drive for Muscularity Scale it was .88.

Self-Objectification Questionnaire. The Self-Objectification Questionnaire (Fredrickson, Roberts, Noll, Quinn, & Twenge, 1998) was used as a check of the self-objectification manipulation. This questionnaire asked participants to rank 10 attributes from most important to their self-concept to least important. Examples of attributes that were ranked were “weight,” “physical attractiveness,”
and “health.” If the self-objectification task was successful, those who were in the self-objectification condition should score higher (i.e., because they have become more appearance focused) on this questionnaire than those in the no self-objectification condition.

Positive and Negative Affect Scale (PANAS). The PANAS (Watson, Clark, & Tellegan, 1988) was used as a check of the self-affirmation manipulation. The PANAS consists of 20 items (e.g., “upset,” “distressed,” “proud”), each rated on a 1 (very slightly or not at all) to 5 (extremely) Likert-type scale, and refers to how participants felt at that moment. Participants who were self-affirmed should report a more positive and less negative affect compared to those who did not undergo the self-affirmation manipulation. Cronbach’s $\alpha$ for the positive affect items was .88, and for negative affect it was .85, and thus they were combined into a positive and a negative affect index, respectively.

State Self-Esteem Questionnaire. The State Self-Esteem Questionnaire (Heatherton & Polivy, 1991) served as an additional check of the self-affirmation manipulation. Examples of items on the State Self-Esteem Questionnaire are, “I feel confident about my abilities” and “I feel displeased with myself.” There were 20 items, each rated on a 1 (not at all) to 5 (extremely) Likert-type scale, and participants were asked to rate how they felt at that moment. Participants who were self-affirmed should have a higher state self-esteem score compared to those who did not undergo the self-affirmation manipulation. Cronbach’s $\alpha$ for the 20 items was .93, and thus they were combined into a total state self-esteem index.

Procedure
After institutional review board approval was given by the University of California San Diego Human Research Protections Program (Protocol #151847S), participants were recruited through the Psychology Department Human Participant Pool online recruitment system and run individually by an experimenter wearing a lab coat that hid her silhouette. The lab coats were meant to reduce the likelihood that participants would compare their bodies to that of their experimenter. To avoid distraction and prevent social media use during the study, participants left their cell phones and laptops in a waiting area. After providing informed consent, participants were randomly assigned to either the self-objectification writing task (designed to focus their attention on their appearance) or a control writing task, each developed by Register et al. (2015). Specifically, in the self-objectification condition, participants were given 5 minutes to “look at yourself from someone else’s perspective. Try to mentally ‘gaze’ at your physical appearance through the eyes of someone else, as if your body were an object to behold. In the space below, explain how this other person sees you and compares your body to the ‘ideal’ body for your gender.” Those participants randomly assigned to the control condition were given 5 minutes to write down all the activities they had participated in during the past 24 hours in chronological order.

Next, participants randomly assigned to the self-affirmation condition completed a 10-minute writing task where they ranked 13 given values (e.g., relations with friends/family, creativity, religion/spirituality) from most important to least important, and wrote about why their most important value is significant to them. This task was based closely on the self-affirmation task designed by Harber (1995) and adapted by Cohen, Aronson, and Steele (2000). However, in the current study, the value “physical attractiveness” was considered to be too closely linked to appearance and was therefore removed. The values of “religion/spirituality” and “kindness” were added to maintain a similar task length and diversify the options available to increase the likelihood that participants would be able to find a “most important” value about which to write. In the control affirmation condition, participants wrote about their least important value, focusing on why this value might be significant to other individuals. In the no affirmation condition, participants did not complete either task and proceeded directly to the Drive for Muscularity and Drive for Thinness questionnaire.1 The self-affirmation task and control task have been used extensively in the self-affirmation literature, and previous research has generally found that self-affirmation effectively boosts self-integrity (Steele, 1988) and mood (Koole et al., 1999).

Participants were then asked to complete The Drive for Thinness subscale and Drive for Muscularity Scale, followed by the Self-Objectification Questionnaire, the PANAS, and the State Self-Esteem Questionnaire in a randomized order. Next,

---

1 A filler task for the no affirmation condition was considered but ultimately not included because such a task might have distracted from the self-objectification manipulation. The benefit of nondistraction outweighed the small 10-minute difference in total experimental task time between the self-affirmation/control affirmation conditions and the no affirmation condition.
participants completed a questionnaire that asked for demographic information (e.g., age, sex, weight, height). Thereafter, all participants regardless of condition completed a different self-affirmation writing task adapted from a study by McGuire and McGuire (1996; Harris, Mayle, Mabbott, & Napper, 2007). This task asked participants to list as many of their strengths and positive qualities that they could think of in 3 minutes. The task was used to reinforce every participant’s self-value in order to combat a possible decrease in state self-esteem and/or mood due to the self-objectification task. Next, a postexperimental inquiry was conducted to discover if participants knew about the purpose of the study. Finally, participants were fully debriefed and thanked for their participation.

Results

Initially, 214 University of California San Diego undergraduates participated. However, seven participants were excluded from data analyses for not following task instructions. Nine participants were excluded because this study depended on an “ideal” image specific to the United States, and these participants spent their formative years (during which time an “ideal” image might be acquired) elsewhere (e.g., India, Philippines). Twenty participants were excluded due to underweight BMI (Body Mass Index) scores (below 18.5) as defined by the Centers for Disease Control and Prevention (CDC). Those who are underweight are more likely to be closer to the “ideal” image promoted by Western society. Therefore, engaging in the self-objectification task would not likely produce the same feelings of not meeting the societal ideal as it would in people who are not already at the ideal weight. The mean Body Mass Index Score for the final sample of 178 participants was 23.20 (SD = 3.55).

Despite these exclusions, a power analysis suggests that analyses involving the main effects of the self-objectification manipulation (reported below) would have sufficient power. Specifically, given that it is the only previous published study that utilized the self-objectification manipulation employed in the present experiment, we utilized the Register et al. (2015) findings on the drive for thinness measure as the best available estimate of the likely effect size for the comparison between the self-objectification and no self-objectification conditions in the present experiment (d = .52). Using Cohen’s (1969) power tables with α set at .05, d at .52, and n at 81 (the lowest n of the two conditions in the present experiment), we found that the best available estimate of our power for the self-objectification main effect was greater than .88.

Unfortunately, given the fact that no previous published work appears to have examined the role of self-affirmation in mitigating the effects of self-objectification, it was not possible to derive an evidenced-based effect size for the self-affirmation analyses. However, if we assume that the true effect size is medium (e.g., .25), then the n of ≥ 58 obtained for each of the three self-objectification conditions in this experiment results in a power greater than .80. Of course, the tests of interaction effects between self-affirmation and self-objectification would have lower power and, particularly for male participants, would be underpowered and should therefore be interpreted with caution.

Manipulation Checks

Self-objectification. A 2 (self-objectification condition: objectified vs. not objectified) x 3 (affirmation condition: self-affirmation vs. control affirmation vs. no affirmation) x 2 (sex: men vs. women) Analysis of Variance (ANOVA) was conducted on participants’ Self-Objectification Questionnaire scores. The results demonstrated a statistically significant main effect of affirmation condition, F(2, 175) = 3.72, p = .026, η² = .04. Specifically, a post-hoc analysis using Fisher’s Least Significant Difference method revealed that those who completed the self-affirmation task (i.e., wrote about their highest ranked value) scored significantly lower on the Self-Objectification Questionnaire compared to those who completed the control affirmation task (i.e., wrote about their lowest ranked value; p = .024) or neither task (i.e., no affirmation condition; p = .006; see Figure 1).

The results also demonstrated a statistically significant main effect of sex, F(1, 176) = 10.96, p < .001, η² = .06. Specifically, women (M = −2.26, SD = 13.05) scored significantly higher on the Self-Objectification Questionnaire compared to men (M = −9.24, SD = 12.08). No other main effects or interaction effects were significant (all p’s > .37; all η² ≤ .06).

Negative and positive affect. Respective 2 (self-objectification condition: objectified vs. not objectified) x 3 (affirmation condition: self-affirmation vs. control affirmation vs. no affirmation) x 2 (sex: men vs. women) ANOVAs were conducted on participants’ negative affect scores and positive affect scores. The results did not demonstrate any differences in negative affect as a function
of condition (all ps > .51; \eta^2 \leq .01). For positive affect, the results demonstrated only a marginal interaction between self-objectification condition and sex, \(F(1,176) = 3.73, p = .055, \eta^2 = .02\). Specifically, men who were not self-objectified reported a more positive affect (\(M = 27.92, SD = 7.12\)) than did women who were in the self-objectification condition (\(M = 26.00, SD = 7.68\)), whereas women who were not self-objectified reported less positive affect (\(M = 24.00, SD = 7.26\)) than did women in the self-objectification condition (\(M = 26.67, SD = 7.62\)).

**State self-esteem.** A 2 (self-objectification condition: objectified vs. not objectified) x 3 (affirmation condition: self-affirmation vs. control affirmation vs. no affirmation) x 2 (sex: men vs. women) ANOVA was conducted on participants’ state self-esteem scores. The results did not demonstrate any main effects or interactions (all ps > .10; all \eta^2 \leq .03).

**Dependent Measures**

**Drive for thinness.** A 2 (self-objectification condition: objectified vs. not objectified) x 3 (affirmation condition: self-affirmation vs. control affirmation vs. no affirmation) x 2 (sex: men vs. women) ANOVA was conducted on participants’ drive for thinness scores. The results demonstrated a statistically significant main effect of affirmation condition, \(F(2, 175) = 3.90, p = .022, \eta^2 = .05\). Specifically, a post-hoc analysis revealed that those who did not complete the affirmation task scored significantly higher on the Drive for Thinness subscale compared to those who wrote about their highest (\(p < .001\)) or lowest (\(p = .049\)) ranked values (see Figure 2). There was no significant difference in drive for thinness scores between those in the self-affirmation or control affirmation conditions (\(p > .16\)).

The results also demonstrated a statistically significant main effect of sex, \(F(1, 176) = 32.71, p < .001, \eta^2 = .17\). Specifically, women (\(M = 3.30, SD = 1.27\)) scored significantly higher than men (\(M = 2.22, SD = 0.94\)) on the Drive for Thinness subscale. No other main effects or interaction effects were significant (all ps > .28; \eta^2 \leq .01). See Table 1 for means and standard deviations of drive for thinness scores as a function of condition.

**Drive for muscularity.** A 2 (self-objectification condition: objectified vs. not objectified) x 3 (affirmation condition: self-affirmation vs. control affirmation vs. no affirmation) x 2 (sex: men vs. women) ANOVA was conducted on participants’ drive for muscularity scores. The results demonstrated a statistically significant main effect of affirmation condition, \(F(2, 175) = 3.47, p = .033, \eta^2 = .04\). Specifically, a post-hoc analysis revealed that those who did not complete the affirmation task scored marginally higher on the Drive for Muscularity Scale than did those who wrote about their highest ranked value (\(p = .077\); see Figure 2).

The results also demonstrated a marginal main effect of self-objectification condition, \(F(1, 176) = 3.11, p = .08, \eta^2 = .02\). Specifically, those who were not self-objectified scored higher on the Drive for Muscularity Scale than did those who were self-objectified.

As was the case with Drive for Thinness, a statistically significant main effect of sex was also found, \(F(1, 176) = 43.29, p < .001, \eta^2 = .21\). Specifically, men (\(M = 2.74, SD = 0.84\)) scored significantly higher than women (\(M = 2.01, SD = 0.63\)) on the Drive for Muscularity Scale.

Additionally, the results demonstrated a marginal interaction between affirmation condition and sex, \(F(2,175) = 2.43, p = .091, \eta^2 = .03\). Specifically,
men who did not complete the affirmation task scored higher on the Drive for Muscularity Scale compared to those who wrote about their lowest or highest ranked value, whereas affirmation condition did not affect women’s scores. No other interaction effects were significant (all \( p \geq .17; \eta^2 \leq .03 \)). See Table 1 for means and standard deviations of drive for muscularity scores as a function of condition.

**Discussion**

We evaluated whether a values-based self-affirmation intervention could remove or reduce the negative effects of self-objectification by first attempting to induce self-objectification and then introducing a self-affirmation intervention. Moreover, we hypothesized that performing a self-objectification task (versus a control task) would result in a higher drive for thinness for women and a higher drive for muscularity for men among those participants who did not also perform a self-affirmation task, but that there would be no difference in drive for thinness or muscularity among those who were self-affirmed. Contrary to our prediction, we did not find any statistically significant main effects of self-objectification or interactions between self-objectification condition and self-affirmation condition on either Drive for Thinness or Drive for Muscularity.

Although the self-objectification manipulation did not produce the significant effects that were predicted, the results of this study are still informative for understanding the role that self-affirmation could have in improving body image. That is, although the self-objectification manipulation does not appear to have altered levels of self-objectification (at least as assessed via The Self-Objectification Questionnaire; Fredrickson et al., 1998), given that participants were randomly assigned to self-affirmation condition, participants across the three self-affirmation conditions should have begun that task with similar levels of self-objectification as well as drive for thinness and muscularity. Thus, any self-affirmation condition differences in self-objectification, drive for thinness, and drive for muscularity scores should be due to the effects of the self-affirmation manipulation. Given that those who had been randomly assigned to complete the self-affirmation task subsequently scored marginally lower on the Drive for Muscularity Scale, as well as the Drive for Thinness subscale, and significantly lower on the Self-Objectification Questionnaire compared to those who did not engage in self-affirmation. Overall, these results suggest that focusing on nonappearance-related values might be a promising strategy for improving general body image.

It is important to note that, although Drive for Thinness, Drive for Muscularity, and Self-Objectification scores were reduced in some way by our affirmation task, it is unknown how long this effect might have lasted. In fact, we would argue that it is unlikely that one 10-minute intervention would be able to keep self-objectification, drive for thinness, and drive for muscularity low, due to the constant exposure to and comparison with media images of seemingly perfect individuals. Future studies might benefit from multiple follow-ups to determine how long the effects of self-affirmation last. From a theoretical standpoint, it will also be important for future work to determine the mechanisms through which self-affirmation may produce beneficial effects on self-objectification and drive for thinness/muscularity. Because previous studies have found that self-affirmation tasks can alter mood and/or improve state self-esteem (Koole et al., 1999; Sherman & Cohen, 2006), we expected that performing our self-affirmation task might remove the effects of self-objectification by decreasing negative affect and/or improving state self-esteem. However, we found no evidence of differences in mood or state self-esteem scores as a function of self-affirmation condition.

There were a few marginal effects involving the self-objectification manipulation that should be briefly mentioned. One of these findings was that those who were self-objectified scored marginally lower on the Drive for Muscularity Scale than
those who were not self-objectified. We suspect that this marginal difference may be a function of particularly female participants (who were the largest proportion of our sample), who had been self-objectified, being less willing to endorse drive for musculature items because musculature is not part of the “thin ideal” in society. The second marginal effect involving the self-objectification manipulation was that women who were not self-objectified reported less positive affect than did women in the self-objectification condition. It is possible that this marginal difference may be due to participants in the no self-objectification condition feeling somewhat less positive as a function of spending 5 minutes listing the activities they had accomplished over the past 24 hours (i.e., perhaps the activities the typical college student accomplishes in a 24-hour period are not particularly positive or inspiring, etc.). It is important to note that both effects were only marginal and neither had been predicted. Thus, the effects should be interpreted very cautiously, and these speculations should be considered highly tentative.

**Sex Differences**

It is also worth discussing the multiple sex differences that emerged in our results, although these should be interpreted with caution due to the small number of men who participated. In our sample, men had a higher drive to be muscular compared to women, and women had a higher drive to be thin compared to men. This provides further evidence that the U.S. female ideal is thin and the U.S. male ideal is muscular (Register et al., 2015). Furthermore, regardless of self-objectification condition, women scored significantly higher on the self-objectification manipulation check than did men. Because a higher score on the Self-Objectification Questionnaire indicates being more appearance focused rather than competence focused, this might suggest that women are naturally more appearance focused. This could be due to the unrelenting media images and advertisements pushing women to obtain beauty (Stice et al., 2000), which is appearance related.

Interestingly, we found that men who were not self-objectified reported a more positive affect than did women who were not self-objectified. This is consistent with the literature on sex differences in affect because women are generally more prone to depression (Nolen-Hoeckema, 2001) and report less positive affect than men (Bojanowska & Zalewska, 2016; Mroczek & Kolarz, 1998).

**Methodological Issues**

This experiment had several methodological strengths including that experimenters were kept blind to condition and wore long lab coats to hide their silhouettes, thereby reducing the likelihood that participants would compare their bodies to that of their experimenter. Additionally, we eliminated distractors, social media, and Internet use during the study by having participants store their belongings (e.g., cell phones and laptops) outside of the study room. According to Derenne and Beresin (2006), mass media plays a significant role in increased body dissatisfaction among men and women. Thus, it was important for us to control for Internet and media use during the study to prevent participants from experiencing additional self-objectification or body dissatisfaction. Given that participants were randomly assigned, it is unlikely that day-to-day social media use differed reliably across conditions. However, it could be beneficial for future work to assess individual differences in social media use and statistically control for any such differences in the analyses. Another methodological strength of this study was that, unlike much of the previous self-objectification literature, we made an effort to experimentally manipulate self-objectification using a task that had been successful in a previous experiment (Register et al., 2015). Unfortunately, the manipulation does not appear to have been successful in this experiment in that those in the self-objectification condition did not score reliably higher on the self-objectification manipulation check than those who were in the no objectification condition (which could explain why there were no differences in drive for thinness and musculature as a function of self-objectification condition). It is difficult to say why the task was not successful in altering self-objectification scores in the present experiment. One possibility concerns the fact that the average BMI for our participants was in the “healthy” range as defined by the CDC, whereas the average BMI for Register et al.’s (2015) study was in the “overweight” range. This might suggest that participants who have lower or “healthier” BMIs are closer to the ideal image promoted by Western society and therefore do not become as easily self-objectified when comparing their bodies to an ideal. Future work might examine this possibility by recruiting participants who are classified as having “healthy” and “overweight” BMIs to directly compare how these two groups are affected by the self-objectification task, as well as how much their body image scores improve with
Self-Affirmation to Remove Self-Objectification

Register et al. (2015) found that their 5-minute self-objectification task designed by Register et al. (2015) has attempted to replicate the effects of the self-objectification manipulation, such as the Objectified Body Consciousness Scale (McKinley & Hyde, 1996).

An additional limitation is the impact of having participants whose ethnic/racial background is not representative of the general population. Although the ethnic breakdown of the sample utilized in our study closely resembles the ethnic makeup of students enrolled at the university where the study was conducted, this does mean the generalizability of our results to the rest of the U.S. population may be a concern. As stated previously, it is possible that a sample with a different distribution of ethnicities may produce different results. Future studies might benefit from recruiting a sample that is more ethnically representative of the United States as a whole. However, an attempt was made to have our sample represent the general “ideals” of the U.S. population by excluding participants who did not spend their formative years in the United States. These participants were excluded because this study depended on an “ideal” image specific to the United States, and these participants spent their formative years (during which time an “ideal” image might be acquired) elsewhere.

Another limitation concerns those participants who were excluded due to underweight BMIs. Of those excluded, only 3 were men and 17 were women. It is possible that the underweight men who were excluded viewed themselves as further away from the ideal, if that ideal was a heavier, more muscular one. Although including these three men in the analyses likely would not have resulted in significantly different results, future studies might benefit from strengthening the self-objectification manipulation and using different/additional measures to assess the efficacy of the self-objectification manipulation check measure used is not sufficiently sensitive to detect differences that remain following such distractions. Future studies may benefit from strengthening the self-objectification manipulation and using different/additional measures to assess the efficacy of the self-objectification manipulation, such as the Objectified Body Consciousness Scale (McKinley & Hyde, 1996).

Another potential reason the task was not successful in altering self-objectification scores in the present experiment concerns the different ethnic breakdown of our sample versus that used by Register et al. (2015). It is possible that a sample with a different distribution of ethnicities may produce different results because different ethnicities could have different body ideals and levels of body image disturbance. For example, the sample used by Register et al. (2015) had a much higher African American (32% vs. 1.7%) and a much lower Asian (15% vs. 41.6%) composition than did our sample. Altabe (1998) found several differences in body image disturbance between different ethnic groups (i.e., White and Hispanic-Americans showed more weight-related body image disturbance than African Americans and Asian Americans). Thus, it is possible that the different distribution in ethnicities was a factor contributing to the differing results between the present study compared to that of Register et al. (2015).

Another potential reason the task was not successful in significantly altering self-objectification scores concerns the possibility that the control self-objectification task focused participants on themselves by having them list their activities over the past 24 hours, inadvertently producing self-objectification scores similar to those in the self-objectification condition. It seems unlikely that a task that requires individuals to simply list their activities would produce self-objectification (i.e., internalization of an observer’s perspective of one’s body). Nevertheless, future studies might benefit from using a self-objectification control task that does not require writing about anything remotely related to the self.

To our knowledge, this is the first study that has attempted to replicate the effects of the self-objectification task designed by Register et al. (2015). In some respects, it is surprising that Register et al. (2015) found that their 5-minute self-objectification writing task actually produced differences on the self-objectification and drive for thinness measures given that those scales assess more enduring values and behaviors. Thus, it is possible that the task is simply not strong enough to produce reliable differences in self-objectification. Relatedly, it should be noted that most participants in the present experiment completed several tasks/questionnaires in between the self-objectification task and the manipulation check. This does not appear to have been the case in the Register et al. (2015) study. Thus, perhaps this particular self-objectification manipulation is not strong enough to withstand such distractions, and/or the particular self-objectification manipulation check measure used is not sufficiently sensitive to detect differences that remain following such distractions. Future studies may benefit from strengthening the self-objectification manipulation and using different/additional measures to assess the efficacy of the self-objectification manipulation, such as the Objectified Body Consciousness Scale (McKinley & Hyde, 1996).

Another possible reason that the task was not successful in altering self-objectification scores in the present experiment concerns the different ethnic breakdown of our sample versus that used by Register et al. (2015). It is possible that a sample with a different distribution of ethnicities may produce different results because different ethnicities could have different body ideals and levels of body image disturbance. For example, the sample used by Register et al. (2015) had a much higher African American (32% vs. 1.7%) and a much lower Asian (15% vs. 41.6%) composition than did our sample. Altabe (1998) found several differences in body image disturbance between different ethnic groups (i.e., White and Hispanic-Americans showed more weight-related body image disturbance than African Americans and Asian Americans). Thus, it is possible that the different distribution in ethnicities was a factor contributing to the differing results between the present study compared to that of Register et al. (2015).

Another potential reason the task was not successful in significantly altering self-objectification scores concerns the possibility that the control self-objectification task focused participants on themselves by having them list their activities over the past 24 hours, inadvertently producing self-objectification scores similar to those in the self-objectification condition. It seems unlikely that a task that requires individuals to simply list their activities would produce self-objectification (i.e., internalization of an observer’s perspective of one’s body). Nevertheless, future studies might benefit from using a self-objectification control task that does not require writing about anything remotely related to the self.
Practical Implications and Conclusions
This study demonstrated that a values-based affirmation intervention has the potential to positively influence body image because writing about nonappearance-related values reduced participants’ drive for thinness, drive for muscularity, and self-objectification. Such a task might be a useful exercise to implement in positive body image programs, which aim to reduce negative feelings about appearance and promote body acceptance. Further, the study also reinforced the idea that women in Western culture prefer a thin ideal image and may be more appearance focused, and men prefer a muscular ideal image and are less appearance focused than women. Future studies implementing follow-ups are needed to determine the longevity and practical significance of the self-affirmation intervention. Furthermore, determining whether a self-affirmation task would also be beneficial for those who are considered overweight would be valuable.

References


Author Note. Sarrah I. Ali, Psychology Department, University of California, San Diego; Heike I. M. Mahler, Psychology Department, University of California, San Diego.

The authors thank Tori Bishop and Lisa Liang for their assistance with data collection. Special thanks to Psi Chi Journal reviewers for their support.

Correspondence concerning this article should be addressed to Heike Mahler, Psychology Department 0109, University of California, San Diego, La Jolla, CA 92093-0109. E-mail: hmahler@ucsd.edu
In the modern world, distraction is unavoidable. Smartphone users face frequent social media notifications and text messages, employees struggle to manage multiple tabs on an Internet browser, music and talk radio is available in even the most desolate locations, and hands-free cell phone and Bluetooth devices allow conversations to continue anytime, anywhere. However, these distractions can be deadly; at any given moment, roughly 660,000 U.S. drivers are using their cellphones while driving (Pickrell & Ye, 2013), and those drivers are 23 times more likely to be involved in a crash or near-crash while on the road than nondistracted drivers (Olson, Hanowski, Hickman, & Bocanegra, 2009). These findings have serious implications in various occupational fields requiring intense concentration in which distraction can produce grave consequences. Those who have yet

The Relationship Between Extraversion and Listening Comprehension Under High- and Low-Salience Visual Distraction Conditions

Nicole Virzi, Steven V. Rouse*, and Cindy Miller-Perrin* Pepperdine University

ABSTRACT. The present study contributed to the body of research examining the link between level of extraversion and response to sensory stimulation. Previous studies have shown that introverts are more susceptible to, and therefore more distracted by, forms of auditory stimulation than extraverts when completing cognitive tasks. However, no study has examined the differing effects of solely visual stimulation on both distraction and cognitive task performance. Using 90 undergraduate college students as participants, this study tested 3 hypotheses: (a) we expected a negative correlation between level of extraversion and self-reported distraction while under high-salience visual stimulation; (b) we predicted a positive correlation between participants’ extraversion score and performance on a listening comprehension task while under high-salience visual stimulation, defined operationally as number of comprehension questions answered correctly; and (c) we expected that the aforementioned correlation would be higher than the correlation between level of extraversion and performance on a listening comprehension task while under low-salience visual stimulation. Although results did not lend support to the idea of these differences in sensory stimulation applying to different forms of visual stimulation (for all correlations, $p = n.s.$), we highlight the theoretical and practical implications of these findings. We provide specific suggestions for future research to help identify those most susceptible to distractions as well as how to best protect individuals from their detrimental consequences.

Open Materials badge earned for transparent research practices. Materials are available at https://osf.io/quy3s/
to enter the workforce are detrimentally affected by distraction as well. Research has shown that the most effective way to improve high school students’ test scores on standardized exams is to enact schoolwide bans on smartphones, which act as sources of both visual and auditory distraction (Beland & Murphy, 2015).

Although multitasking is often expected in many different situations, not all individuals are able to handle outside distractors easily; some find it difficult to concentrate even in the most tranquil environments, and others cannot focus on a task at hand without being overwhelmed by the hum of hectic daily life. When it comes to susceptibility to outside distraction, it appears that individual differences matter more than the actual distractors themselves, and research on sensory distraction has shown that these crucial differences lie in individual personality traits (Eysenck & Graydon, 1989; Furnham & Allass, 1999; Furnham & Bradley, 1997). Knowing exactly which personality characteristics make individuals more susceptible to outside distraction has practical applications in vehicles, workplaces, school, and other social settings. For instance, knowing that drivers with extreme scores on certain dimensional personality characteristics are more likely to be distracted by conversing with a passenger while driving can have a strong preventative impact on accident and mortality rates through civilian education. Similarly, a teacher who understands why seating students next to a window will have a differential effect on individual performance in class based on respective personality characteristics has the opportunity to amplify the chance of success for each and every student, sparking a positive ripple effect extending far beyond the confines of a classroom.

**Theories of Stimulation and Performance**

The Yerkes-Dodson law (Yerkes & Dodson, 1908) states that there is a distinct bell-shaped relationship between arousal and task performance, with increasing levels of arousal being associated with better task performance up to a point, after which higher levels of arousal lead to a decrease in performance. People therefore have an arousal level at which they perform best, although this “sweet spot” is unique to each individual. Hans Eysenck attributed this variance in people’s ability to handle outside distractors to their level of extraversion and posited that the basis of this personality difference was physiological in nature. Eysenck suggested that the optimal level of stimulation was lower for introverts than for extraverts, and that this difference was determined by the stringency of the ascending reticular activating system (ARAS), which is connected to the cerebral cortex and is responsible for filtering the flow of outside stimulation to the brain (Eysenck, 1967). Eysenck’s theory proposed that the ARAS of introverts was not as stringent as the ARAS of extraverts, allowing a surplus of stimulation to reach the brain and causing introverts to reach their peak optimal level of arousal more quickly than extraverts. Conversely, the ARAS of extraverts is highly stringent, filtering out more outside stimulation and therefore leading extraverts to crave social and arousing environments in order to reach their optimal level of arousal.

**Early Sensory Studies**

Ensuing research tested the theory that, when faced with equal levels of external stimulation, the way individuals respond is dependent upon their level of extraversion. Corcoran’s (1964) research created a ripple effect when this theory was applied to the sense of taste and demonstrated that introverts salivated more than extraverts when drops of lemon juice were placed on their tongues. Eysenck replicated these results in adults (Eysenck & Eysenck, 1967), and later research confirmed these findings among college students (Howarth & Skinner, 1969) and children (Casey & McManis, 1971). Complementary studies extended this lower sensory threshold to pain perception as well, finding a positive correlation between level of extraversion and pain tolerance (Haslam, 1967). These studies supported Eysenck’s theory as applied to sensory stimulation.

It has also been shown that introverts and extraverts experience these cortical arousal differences during routine activities; when given the choice of study location in a library, there is a positive relationship between college students’ level of extraversion and the amount of potential distraction in their preferred location (Campbell & Hawley, 1982). Test scores covering retention of studied material, however, have not been compared.

**Auditory and Musical Distraction**

Eysenck’s interpretation of the Yerkes-Dodson law has most commonly been demonstrated through studies of auditory stimulation. When individuals scoring in the extremes on the Eysenck Personality Inventory (EPI) Extraversion scale were presented with a paired-associate learning task and had the
ability to adjust the intensity of white noise distraction, the mean volume for introverts was much lower than that of extraverts (Geen, 1984). However, physiological measures indicated that both groups were equally aroused during their choice volume exposure. Furthermore, when introverts were assigned to complete the learning task while listening to preferred extravert white noise levels, their performance dropped markedly.

Results of past research have also revealed that, when noise level is increased from quiet (45dB) to low-level noise (60dB), introverts are both more physiologically aroused and perform worse on Law School Admissions Test (LSAT) reading comprehension tests than their extraverted counterparts (Standing, Lynn, & Moxness, 1990). Interestingly, one study concluded that the relationship seems to be affected by the task-relevance of the noise; when a high-complexity problem was paired with task-relevant noise, introverts performed much worse than extraverts under the same conditions. However, when the distraction stimuli were not related to the task stimuli, both groups performed equally well. Being that this was the only study to consider stimulation relevance and results have not been replicated, no definitive conclusions can be drawn (Eysenck & Graydon, 1989).

Auditory stimulation frequently exists in the form of music, which has been shown to have a more complex impact on distraction. For introverts, both short term memory and reading comprehension abilities suffer when completed in the presence of music compared to baseline scores taken during silence, but scores for extraverts do not differ between conditions. Moreover, the level of distraction reported by participants in a posttest questionnaire correlates negatively with Eysenck Personality Questionnaire (EPQ) Extraversion scores, suggesting that introverts find the same level of music to be more distracting than their extraverted counterparts (Furnham & Bradley, 1997). Interestingly, other factors found to correlate positively with EPQ scores include reported frequency of radio listening while working as well as the frequency of radio listening in general.

As noted with noise distractions, the complexity of the music can impact individual cognitive abilities as well. Research has found significant interactions between an individual’s level of extraversion and performance on both memory and observation tests in the presence of both simple and complex music, categorized by variations in “tempo, rhythmic tonality, melodic complexity and vocal meaningfulness” (Furnham & Allass, 1999). Interestingly, results have noted a crossover effect: when exposed to simple music, introverts perform better on cognitive tasks when compared to baseline scores during silence, but extraverts perform worse. When exposed to complex music, extraverts perform better on cognitive tasks when compared to baseline scores during silence, and introverts perform worse. Although the same directional trends were found during assessment of reading comprehension tasks, the results were not significant. No difference was reported in distraction between introverts and extraverts during the simple noise condition, and a large discrepancy was found during the complex music task.

**Television Distraction**

Accordingly, outside stimulation involving both auditory and visual distraction has proven to be just as impactful. When completing Graduate Management Admission Test (GMAT) reading comprehension passages in front of a television playing a popular drama series, introverts performed significantly worse on passage questions than extraverts, with no difference in proficiency existing between the two groups during silence (Furnham, Gunter, & Peterson, 1994). Research has also shown that the addition of television distraction during the question portion of reading comprehension assessment presents no further decrement in reading comprehension ability for either personality type, supporting the idea that forms of distraction impact cognitive abilities during encoding of information as opposed to retrieval (Armstrong & Chung, 2000).

It is clear that differences between individuals extend beyond their levels of extraversion, yet extraversion is the only Big Five trait that has shown any impact on reading comprehension while in the presence of television distraction (Ylias & Heaven, 2003). Neuroticism, openness, agreeableness, and conscientiousness have thus far failed to show any involvement in the complex relationship between reading comprehension and both auditory and visual stimulation.

**Present Study**

Although the relationships between auditory distraction, reading comprehension, and extraversion have been studied at length, very limited research has examined the impact of sources of distraction that are purely visual. A handful of studies dating back to the early and mid-20th century that attempted to answer this
Extraversion and Listening Comprehension

The current study examined three hypotheses, the first being that there would be a negative correlation between level of extraversion and self-reported distraction while under high-salience visual stimulation—meaning that the higher a participant's level of extraversion, the lower their reported level of distraction as a result of viewing high-salience visual stimulation. Data supporting this first hypothesis would be the first to suggest that Eysenck's theory of sensory threshold differences between introverts and extraverts—shown by previous research to be applicable to the senses of taste (Corcoran, 1964), touch (Haslam, 1967), and hearing (Furnham & Bradley, 1997; Geen, 1984)—could be extended to isolated visual stimulation as well. Since Eysenck's theory utilized the term “sensory” as an all-encompassing label, results of this study assessing visual stimulation should mirror the promising results of studies that assessed other forms of sensory stimulation.

Second, we hypothesized a positive correlation between participants' extraversion score and performance on a listening comprehension task (defined operationally as the number of comprehension questions answered correctly) while under high-salience visual stimulation. This would indicate that the higher a participant's level of extraversion, the more comprehension questions the participant would answer correctly for passages listened to while under high-salience visual stimulation. Data supporting the second hypothesis would demonstrate that the differential relationships between auditory stimulation salience and comprehension performance between introverts and extraverts (as supported by previous studies) can also be elicited from visual stimulation, further affirming the inclusive nature of Eysenck's theory of sensory stimulation.

Finally, we hypothesized that the correlation posited in our second hypothesis would be stronger than the correlation between level of extraversion and performance on a listening comprehension task while under low-salience visual stimulation. Confirmation of our third hypothesis would suggest that the more salient the stimuli, the better extraverted individuals perform—even if the form of stimuli is visual. In turn, this would also be the first study to demonstrate that the impact of differing levels of auditory stimulation salience on comprehension performance (Furnham & Allass, 1999; Furnham & Bradley, 1997) can also be elicited from differing levels of visual stimulation, providing further support for the idea that Eysenck's theory of sensory stimulation and performance can be applied to any sensory modality.

Method

Participants

This study included 111 undergraduates, 20 of whom were excluded from analyses because they obtained test scores suggesting invalid response styles (as described in the next paragraph) and one of whom became ill and excused himself during a data collection session. The final dataset consisted of 90 participants (47 women and 43 men; \( M_{\text{age}} = 19.00, SD = 1.13 \)) enrolled in foundation psychology courses during fall 2016. Self-reported ethnicity was as follows: 56.7% of participants identified as White/European American, 22.2% as Asian or Asian-American, 8.9% as Hispanic/Latino, 4.4% as Black/African American, 1.1% as American Indian/Alaska Native, and 1.1% as Native Hawaiian/Other Pacific Islander. Participant socioeconomic status while growing up was as follows: 11.1% reported belonging to the upper class, 56.7% to the upper-middle class, 22.2% to the middle class, 5.6% to the lower middle class, and 4.4% to the lower class. Participants were recruited through an online research management system. Participants received research credits as part of a requirement for successful completion of the course.

Materials

Eysenck Personality Inventory (EPI). Each participant completed the EPI, a 57-item scale that assesses levels of both extraversion and neuroticism by having participants respond to statements with either a “yes” or a “no.” The EPI also includes a 9-question Lie scale. Reliability coefficients ranging from .50 to .87 have been reported (Farley, 1971). For the purpose of this study, only the 24-item Extraversion scale was included in analyses, although abnormally high scores on the Lie scale excluded a total of 20 participants from the sample.

Listening comprehension. Passages and questions were selected from Peterson’s Master the Catholic High School Entrance Exams 2014 booklet (“Reading Comprehension,” 2013). Passages were selected based on length, number of questions, and diversity of topic. The text of the passages was

Materials

Eysenck Personality Inventory (EPI). Each participant completed the EPI, a 57-item scale that assesses levels of both extraversion and neuroticism by having participants respond to statements with either a “yes” or a “no.” The EPI also includes a 9-question Lie scale. Reliability coefficients ranging from .50 to .87 have been reported (Farley, 1971). For the purpose of this study, only the 24-item Extraversion scale was included in analyses, although abnormally high scores on the Lie scale excluded a total of 20 participants from the sample.

Listening comprehension. Passages and questions were selected from Peterson’s Master the Catholic High School Entrance Exams 2014 booklet (“Reading Comprehension,” 2013). Passages were selected based on length, number of questions, and diversity of topic. The text of the passages was
presented orally to participants through a prerecording to ensure standardization between groups. Each passage was followed by eight multiple-choice questions with answer choices A through D that assessed participants’ comprehension of the passage. The listening comprehension text and passage questions have been archived at https://osf.io/quy3s/.

**Distraction questionnaire.** Participants completed a single-item questionnaire asking them to rate how distracted they were during each condition using a response scale that ranged from 1 (*not at all distracted*) to 10 (*very distracted*). The distraction questionnaire has been archived at https://osf.io/quy3s/.

**Low-salience visual stimulation.** A 5-minute muted sample of a video of crashing beach waves was used to create relatively low-salience visual distraction. Due to the predictable and repetitive nature of crashing beach waves, researchers and collaborators agreed that the scene provided participants with a form of low-complexity stimulation that would incite levels of objective stimulation comparable to—as well as serve as the rough visual equivalent of—the low-level noise (Standing et al., 1990) and low-complexity music (Furnham & Allass, 1999) employed by researchers studying auditory forms of distraction. The video has been archived at https://osf.io/quy3s/.

**High-salience visual stimulation.** A 5-minute muted sample of a *Looney Toons* cartoon was used to create high-salience visual stimulation. Due to the lack of predictable visual content, speed of character movement, and erratic nature of the plotline of the cartoon, researchers and collaborators agreed that the scene provided participants with a visual form of high-complexity stimulation that would incite relatively high levels of distraction among participants in the same way that high-level noise (Standing et al., 1990) and high-complexity music (Furnham & Allass, 1999) was able to do in previously performed auditory distraction studies. The video has been archived at https://osf.io/quy3s/.

**Design**

This study was designed as a within-subjects experiment. The independent variable was the type of visual stimulation and contained three levels: the control condition (including no visual stimulation), low-salience visual stimulation, and high-salience visual stimulation. The dependent variables were auditory comprehension performance—defined operationally as the number of passage questions answered correctly by the participant—and self-reported distraction ratings between the different conditions. Data collection took place over four different experimental sessions, each including between 16 and 25 participants who were then assigned to two different classrooms.

**Procedure**

After institutional review board approval (16-04-260), participants signed up for the study on an online research management system and provided informed consent. One-hour research sessions were conducted in classrooms. Upon arrival, participants were given prenumbered optical-scanned answer sheets that randomly assigned them to proceed to one of two classrooms. In the first classroom, participants completed the control condition task, the low-salience visual stimulation task, and then the high-salience visual stimulation task. This classroom was under the direction of the principal investigator. To counterbalance conditions and rule out the potential for any order effects, participants assigned to the second classroom completed the control condition task, the high-salience visual stimulation task, and then the low-salience visual stimulation task. This classroom was under the direction of a research assistant. The number of participants in each experimental session were split evenly between the two classrooms.

Once all participants were seated, each received a copy of the EPI and was asked to answer the questions as honestly and accurately as possible. Researchers told participants to wait quietly upon completion for further instructions.

Once all participants were finished, researchers collected the EPI assessments and set up the blank projector screen to begin the control condition. Researchers told participants that they would be listening to a recording of somebody reading a comprehension passage aloud and would be asked to answer questions based on the passage when the recording was completed. Researchers directed participants to keep their auditory attention on the passage and their visual attention on the blank projector screen.

The recorded passage was played for 5 minutes, and then researchers distributed the listening comprehension questions. Researchers advised participants to answer the questions to the best of their ability and record their answers on their answer sheet. Researchers told participants they had 5 minutes to answer the questions and to sit quietly once they had completed the questions.
In the first classroom, participants completed two more auditory comprehension tasks, the first while under the influence of the low-salience visual stimuli and the second while under the influence of the high-salience visual stimuli. In the second classroom, participants listened to the same passages in the same order, but the visual stimulations were presented in reverse order: the first experimental passage was presented with the high-salience stimuli (cartoon video) while the second passage was presented with the low-salience stimuli (waves video).

Once both experimental conditions were completed, participants were given the Distraction Questionnaire and rated their level of distraction during each of the three conditions. Questionnaires and answer sheets were then collected and the researchers thanked the students for their participation in the study. After all research sessions were completed, all participants were entered into a raffle to win a $100 Amazon gift card.

**Results**

**EPI Results**
Scores on both the Extraversion scale (M = 15.31, SD = 4.36) and the Neuroticism scale (M = 14.04, SD = 4.44) were found to be normally distributed and internally consistent for the sample (for both scales, α = .79). Mean scores on the Extraversion scale between participants assigned to Classroom 1 (M = 15.32, SD = 4.24) and Classroom 2 (M = 15.18, SD = 4.29) presented no significant difference between groups (p = .88, d = .03).

**Listening Comprehension Passages**
Scores on the control passage (M = 5.79, SD = 1.43), first experimental passage (M = 4.71, SD = 1.96), and second experimental passage (M = 4.38, SD = 1.63) were found to be normally distributed but not internally consistent. Cronbach’s αs were .37, .61, and .47, respectively.

**Independent-Samples t Tests**
As shown in Table 1, the participants in Classroom 1 and Classroom 2 differed significantly on the number of listening comprehension passage questions answered correctly, with the mean scores of participants in Classroom 1 (M = 5.46, SD = 1.57) being significantly lower than the mean scores of those in Classroom 2 (M = 6.14, SD = 1.19), t(88) = -2.30, p = .02, 95% CI [-1.27, -.09]. Cohen’s effect size value (d = .49) suggested a moderate practical significance. Additionally, a significant difference in mean self-reported distraction rating during the low-salience visual stimulation condition was found between the participants in Classroom 1 (M = 5.17, SD = 2.22) and Classroom 2 (M = 6.44, SD = 2.40), t(88) = -2.66, p = .01, 95% CI [-1.23, -.32]. Cohen’s effect size value (d = .55) suggested a moderate practical significance. There were no other significant differences in dependent variable means between Classroom 1 and Classroom 2.

**Paired-Samples t Tests**
As shown in Table 2, there was not a significant difference in mean overall comprehension scores between the low-salience and high-salience distraction conditions (M = 4.67, SD = 1.75 and M = 4.51, SD = 1.76, respectively), t(88) = .75, p = .46, 95% CI [-.28, .62]. A significant difference was found between the means of self-reported distraction in the low-salience conditions (M = 5.79, SD = .25) and high-salience conditions (M = 7.65, SD = 1.87), t(88) = -6.81, p = .00, 95% CI [-2.40, -1.32].

**Correlations**
Pearson product-moment correlations were used to examine the relationship between the following continuous variables: extraversion scores, self-reported distraction scores, and listening comprehension scores. As shown in Table 3, correlation coefficients measuring the relationship between self-reported distraction scores during control, low-salience and high-salience conditions and extraversion scores were valued between -.05 and .03 (for all coefficients, p = n.s.). Coefficients measuring the relationship between listening comprehension scores for each condition and extraversion scores were valued between -.10 and -.03 (for all coefficients, p = n.s.). The correlation between extraversion scores and listening comprehension scores while under low-salience visual stimulation was -.10, while the correlation between extraversion scores and listening comprehension scores while under high-salience visual stimulation was -.09.

**Discussion**
The purpose of this study was to test the idea that Hans Eysenck’s theory of differential optimal arousal—and the significant differences shown by past research to exist between introverts and extraverts for auditory, tactile, and gustatory stimulation—could also be generalized to visual stimulation. Further, this study sought to determine whether the results of previous research revealing significant differences in performance on cognitive
tasks while in the presence of auditory stimulation between introverts and extraverts could be replicated under the influence of visual stimulation.

The results of the current study did not support any of the three hypotheses. The correlation between participants’ extraversion scores and self-reported distraction under the high-salience visual distraction (cartoon video) condition was not significant. Likewise, the correlation between participants’ extraversion score and performance on the listening comprehension task under the high-salience visual distraction condition was not significant, nor was this correlation significantly higher than the correlation between participants’ extraversion score and performance on the listening comprehension task while under the low-salience visual distraction (waves) condition. This was the first published systematic study to extend Eysenck’s theory of differential optimal arousal to the sense of sight. However, results of this study did not support the idea that these significant differences could also be elicited by visual stimulation.

The results of the current study can be interpreted as lending support to one of two conclusions. The first is that differences in sensory stimulation thresholds and distraction tolerance between introverts and extraverts may be limited only to forms of auditory, gustatory, and somatosensory stimulation and are not applicable to visual stimulation, in which case Eysenck’s theory of sensory threshold differences is not applicable to all bodily senses and is therefore in need of revision. This conclusion would also have practical implications, encouraging public efforts aimed at reducing auditory distraction to focus primarily on forms of distracting stimuli that target the auditory senses. After completion of the study, however, researchers became aware of pitfalls in our research design as well as various ways in which we could have improved upon our study methodology. This introduced an alternate conclusion: this study—being the first that we are aware of to test the effect of sources of visual stimulation—can be interpreted as a critical pilot study with limitations.

One of the most substantial issues that arose in this study was the low internal consistency for each of the three reading comprehension scores (as made evident by all Cronbach’s α scores falling below .62) which suggests that the reading comprehension scores were psychometrically problematic. Previous research examining the impact of auditory distraction upon reading comprehension abilities utilized passages from published standardized tests such as the Law School Admissions Test (Standing et al., 1990) and Graduate Management Admissions Test (Furnham et al., 1994). However, passages from these tests were determined to be too complex to be translated into listening comprehension passages, with many questions requiring readers to refer

<table>
<thead>
<tr>
<th>TABLE 1</th>
<th>Differences in Group Means for Distraction, Comprehension, and Extraversion Scores</th>
</tr>
</thead>
<tbody>
<tr>
<td>Classroom 1 (n = 47)</td>
<td>Classroom 2 (n = 43)</td>
</tr>
<tr>
<td><strong>Self-reported distraction</strong></td>
<td></td>
</tr>
<tr>
<td>Control</td>
<td>3.47 (2.36)</td>
</tr>
<tr>
<td>Waves</td>
<td>5.17 (2.22)</td>
</tr>
<tr>
<td>Cartoon</td>
<td>7.70 (1.97)</td>
</tr>
<tr>
<td><strong>Comprehension scores</strong></td>
<td></td>
</tr>
<tr>
<td>Control</td>
<td>5.46 (1.57)</td>
</tr>
<tr>
<td>Waves</td>
<td>4.85 (1.99)</td>
</tr>
<tr>
<td>Cartoon</td>
<td>4.26 (1.87)</td>
</tr>
<tr>
<td>Extraversion score</td>
<td>15.32 (4.24)</td>
</tr>
</tbody>
</table>

Note. Ranges of possible values for variables of interest: self-reported distraction (1–10), comprehension scores (1–8), extraversion score (0–24).

<table>
<thead>
<tr>
<th>TABLE 2</th>
<th>Differences in Sample Means for Comprehension and Distraction Scores</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total sample (n = 90)</td>
<td></td>
</tr>
<tr>
<td><strong>Waves condition</strong></td>
<td><strong>Cartoon condition</strong></td>
</tr>
<tr>
<td>Comprehension scores</td>
<td>4.67 (1.75)</td>
</tr>
<tr>
<td>Self-reported distraction</td>
<td>5.79 (0.25)</td>
</tr>
</tbody>
</table>

Note. Ranges of possible values for variables of interest: self-reported distraction (1–10), comprehension scores (1–8).

<table>
<thead>
<tr>
<th>TABLE 3</th>
<th>Correlations: Extraversion Scores, Self-Reported Distraction, and Listening Comprehension</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>r Correlation with Extraversion Scores (n = 90)</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Self-reported distraction</strong></td>
<td></td>
</tr>
<tr>
<td>Control</td>
<td>.00</td>
</tr>
<tr>
<td>Waves</td>
<td>.03</td>
</tr>
<tr>
<td>Cartoon</td>
<td>-.05</td>
</tr>
<tr>
<td><strong>Comprehension scores</strong></td>
<td></td>
</tr>
<tr>
<td>Control</td>
<td>-.03</td>
</tr>
<tr>
<td>Waves</td>
<td>-.10</td>
</tr>
<tr>
<td>Cartoon</td>
<td>-.09</td>
</tr>
</tbody>
</table>
Extraversion and Listening Comprehension | Virzi, Rouse, and Miller-Perrin

back to specific lines or terms in the passage—an impossible task when individuals are only listening to a passage and do not have a printed version of the text at their disposal.

The passages and questions chosen for this study were taken from a published practice test booklet for high school entrance exams; the passages were seen as having low levels of complexity and the accompanying questions were believed to test only a surface-level comprehension not requiring repeated reference to passage content. The practice tests were readily accessible to researchers, unlike many official high school entrance exams which are not released to the public. However, the low internal reliability for each of the scores suggests that the assessments were questionable measures of listening comprehension abilities. As a result, it is difficult to know whether the nonsignificant findings were due to an actual absence of a relationship between level of extraversion and listening comprehension under the experimental conditions, or to a statistical artifact associated with the low internal consistency of the comprehension scores. Future research should utilize standardized, published, and psychometrically strong measures to test listening comprehension abilities. Passages and accompanying questions could be taken from the Oral Passage Understanding Scale (OPUS™), which has demonstrated robust internal consistency, test-retest reliability, and interrater reliability (Carrow-Woolfolk & Klein, 2017).

In addition, although randomly assigning participants to one of the two classrooms to counterbalance effects certainly strengthened our experimental methodology, we did find significant differences in dependent variables of interest between classroom groups. Analysis of control condition listening comprehension scores revealed that there was a significant difference in baseline listening comprehension abilities between the groups despite the use of standardized procedures, passage recordings, and scripts/instructions. There was also a significant difference between classrooms in self-reported distraction level elicited by the low-salience visual stimulation condition.

Although every effort was made to create identical experiences for participants in each classroom by standardizing variables over which we had control, we had to secure available classrooms for experimental sessions through the university’s administration, and due to class schedules and concurrent classroom availability, we were given two very different types of classrooms in which to carry out our research. Classroom 1 was a large auditorium-style lecture hall with stadium seating, and Classroom 2 was a smaller classroom with movable seats and desks suitable for discussion-based classes. We also realized that our decision to separate participants into two groups automatically allowed for naturalistic differences to arise between the settings: for example, frequent sneezing by a sick participant in one classroom could have had a large impact on the experience of surrounding participants. There also could have been subtler differences between the classrooms (e.g., lighting, seating arrangement, noise level produced by the movement of chairs) that could have contributed to the creation of very different classroom atmospheres.

Though these classroom differences were limitations of our study, it is unlikely that they accounted for the nonsignificant findings. Because we utilized a within-subjects design, and all three of our hypotheses required only intra-participant data, differences between classrooms presented no major issue if those differences remained consistent throughout the experimental sessions. However, to strengthen methodology, future research should take measures to ensure that, if participants are separated to counterbalance order effects, researchers exercise better control over condition environment to create similarity when at all possible.

Further, because no previous published studies have tested the impact of a solely visual stimulus, we had no precedent to guide either the experimental stimuli selected or the best way in which to elicit different salience levels of visual stimuli. As mentioned in our Method section, we utilized our best judgment in determining what would be seen as low- and high-level visual salience, and the mean self-reported distraction ratings of both classrooms as listed in Table 1 lend support to the idea that we succeeded in eliciting differential levels of distraction among the conditions. However, we neglected to collect feedback regarding the distracting qualities of the videos before executing our study, and this would have been an effective way to lend support to our choices. Doing this also might have encouraged us to find a source of high-salience visual distraction more powerful than the cartoon video to include in our study. Follow-up studies should test the differential distracting qualities of visual stimuli before deciding which stimuli to use in order to confirm that the stimuli will be effective in eliciting targeted levels of visual distraction. In turn, this could help establish a set of standardized sources of visual stimulation for use in future
work environments to maximize success amongst dangerous and frequently fatal consequences of distraction as well as understanding what types of individuals most likely to be impacted by sources of distraction. The more adept society will be at both identifying characteristics and power of sources of distraction, the greater the likelihood that human error will be eliminated between individual differences in personality types and sensory thresholds, as well as help to inform policy efforts to manipulate personality types and sensory thresholds, to clarify and refine the proposed theory between current study be revisited in future research to help understand differences seen to exist between introverts and extroverts. The applicability and suggest that the sensory threshold for the idea that this theory can be applied to the research methodology are able to provide support controlled experimental conditions and different procedures to increase sample size and explore whether significant results are obtained when this procedure is replicated for extreme groups.

In conclusion, the present findings did not support our hypotheses. If future studies continue to obtain nonsignificant results, and Eysenck’s theory of extraversion and sensory thresholds does not appear to apply to visual stimulation, the theory may require revision. However, if more controlled experimental conditions and different research methodology are able to provide support for the idea that this theory can be applied to the visual senses as well, this would affirm the theory’s applicability and suggest that the sensory threshold differences seen to exist between introverts and extraverts extends to visual stimulation.

It is crucial that the hypotheses posited in the current study be revisited in future research to help clarify and refine the proposed theory between personality types and sensory thresholds, as well as help to inform policy efforts to manipulate distractions where they matter most: in vehicles, workplaces, schools, and other social settings. The more research that expands upon the relationship between individual differences in personality characteristics and power of sources of distraction, the more adept society will be at both identifying individuals most likely to be impacted by sources of distraction as well as understanding what types of stimulation can affect individuals most. Together, these benefits can contribute to the prevention of dangerous and frequently fatal consequences that often arise as a direct result of distraction while also informing efforts to curtail school and work environments to maximize success amongst heterogeneous groups of students and employees.

References


Virzi, Rouse, and Miller-Perrin | Extraversion and Listening Comprehension

SUMMER 2018

PSI CHI JOURNAL OF PSYCHOLOGICAL RESEARCH

COPYRIGHT 2018 BY PSI CHI, THE INTERNATIONAL HONOR SOCIETY IN PSYCHOLOGY (VOL. 23, NO. 3/ISSN 2325-7342) 217


Author Note. Nicole Virzi, Social Sciences Division, Pepperdine University; Steven V. Rouse, https://orcid.org/0000-0002-1080-5502, Social Sciences Division, Pepperdine University; Cindy Miller-Perrin, Social Sciences Division, Pepperdine University.

The authors would like to thank Pepperdine University for its funding of this research as well as Alexis Williams for her outstanding assistance. Special thanks also to Psi Chi *Journal* reviewers for their support.

Correspondence concerning this article should be addressed to Nicole Virzi, 6 Iron Latch West, Upper Saddle River, NJ 07458. E-mail: Nicole.Virzi@gmail.com
A prominent goal for individuals who engage in disordered eating behaviours is weight loss, which is often pursued through caloric restriction. The cognitive and behavioural processes that underlie disordered eating, however, remain poorly understood. One aspect that complicates research in this domain is the apparent dissociation between implicit and explicit attitudes toward food in disordered and nondisordered populations. Explicit attitudes refer to consciously accessible mental inclinations and implicit attitudes constitute remnants of past experience, inaccessible to conscious reflection, that influence positive or negative thoughts, actions, or behaviours toward a target (Nosek, Hawkins, & Frazier, 2012). Implicit and explicit attitudes are distinct constructs but may be congruent for an individual depending upon the target of the attitudes and individual differences. Conversely, explicit and implicit attitudes toward a single target may be in conflict (Nosek et al., 2012). For example, Czyzewska and Graham (2008) observed negative implicit attitudes and positive explicit attitudes toward low-calorie food in the same group of individuals. Explicit attitudes can be influenced by a number of factors including intentions, normative attitudes, and social desirability. As such, attitudes assessed through explicit means may not accurately reflect cognitions. Further, explicit attitudes may not be predictive of behaviour, particularly in the realm of eating behaviour. Multiple lines of research suggest the utility of studying implicit processes with regard to eating behaviour. Decisions surrounding food selections are strongly influenced by implicit attitudes and other forms of automaticity, even when these attitudes contradict explicit attitudes (Cervellon, Dubé, & Knäuper, 2007; Yang et al., 2012).

**ABSTRACT.** The purpose of the present study was to investigate the relationship between disordered eating and implicit attitude toward the caloric value of food and, furthermore, to assess whether the personality dimension of reward responsiveness or the more specific construct of reward-based eating drive could better account for contradictory findings in the literature. University student volunteers (N = 100) completed an online questionnaire battery before attending a laboratory session and completing an implicit association test assessing the differential evaluation of high- and low-calorie food. A positive implicit attitude toward low-calorie food was observed in a large proportion of participants (94%). Reward responsiveness was found to moderate the relationship between disordered eating and implicit caloric-related attitude, 95% CIs [0.02, 0.07]. Among those high in reward responsiveness, disordered eating predicted a stronger positive implicit attitude toward low-calorie food. Reward-based eating drive did not moderate the association between disordered eating and implicit caloric-related attitude, 95% CIs [-0.10, 0.14]. The obtained results support the idea of approach and avoidance temperaments, characterized by sensitivity to reward and punishment, and offer evidence of an eating-related behavioural manifestation of such temperaments.

**Reward Responsiveness Moderates Individuals With Disordered Eating's Implicit Attitudes Toward the Caloric Value of Food**

Brittany A. Mascioli and Ron Davis*
Lakehead University

*Faculty mentor
Research surrounding implicit attitudes and eating behaviour has produced mixed results. Using an affective priming task, which is a cognitive method used to measure implicit attitudes, Czyzewska and Graham (2008) found a positive implicit attitude toward high-calorie sweet food and a negative implicit attitude toward both low-calorie food and high-calorie nonsweet food. By contrast, using a unipolar implicit association test (IAT) variant that enables positive and negative implicit attitudes to be assessed separately; more specifically, Houben, Roefs, and Jansen (2010) found a negative implicit attitude toward high-calorie food relative to low-calorie food. When high- and low-calorie food were examined separately, however, a positive implicit attitude toward high-calorie food emerged (Houben et al., 2010). Using the traditional bipolar IAT, Mai, Hoffmann, Hoppert, Schwarz, and Rohm (2015) found a positive implicit attitude toward diet food including reduced-sugar and reduced-fat food items.

Implicit measures have also revealed mixed findings regarding the attitudes of individuals toward the caloric value of food in the restrained eating population. Dietary restraint, defined as the consistent and intentional limitation of food consumption, has been shown to be associated with a positive implicit attitude toward both low-calorie (Maison, Greenwald, & Bruin, 2001) and high-calorie food (Hoefling & Strack, 2008; Houben et al., 2010). Taken together, these findings suggest that the nature of implicit attitudes toward food among individuals with disordered eating may be contingent upon a moderating variable.

The present study aimed to test two potential moderators of the relationship between disordered eating and the implicit attitude toward the caloric value of food. Reward responsiveness, characterized by one’s sensitivity to rewarding stimuli, and reward-based eating drive, which describes a tendency toward uninhibited food consumption, were each assessed as putative moderators of the predictive ability of disordered eating regarding implicit attitude toward the caloric value of food (Epel et al., 2014; Van den Berg, Franken, & Muris, 2010).

One personality construct that may moderate the implicit attitude of individuals with disordered eating toward the caloric value of food is approach-avoidance motivation. This refers to the coordination of behaviour either toward rewarding stimuli or away from threatening or punishing stimuli. Two personality temperaments have been proposed on the basis of the idiosyncratic nature of sensitivity to reward and punishment (Elliot & Thrash, 2002, 2010). Approach temperament is characterized by attentiveness to cues of potential reward and behavioural orientation toward positive stimuli. It comprises high reward responsiveness and low punishment sensitivity. Avoidance temperament is often described as the attentiveness to cues of potential threat and behavioural orientation away from negative stimuli. It is comprised of low reward responsiveness and high punishment sensitivity (Elliot & Thrash, 2002). Inconsistent findings regarding sensitivity to reward and punishment have emerged in the disordered eating population; both low (Harrison, Sternheim, O’Hara, Oldershaw, & Schmidt, 2016) and high (Glashouwer, Bloot, Veenstra, Franken, & de Jong, 2014) reward responsiveness have been observed in anorexia nervosa. One possibility to the inconsistency observed—with regard to reward responsiveness and implicit attitudes toward the caloric value of food among individuals with disordered eating—is that these two constructs covary in a systematic way in this population, thus creating two distinct personality subgroups, each comprising a unique behavioral strategy in line with the common goal of weight control. Individuals with disordered eating of approach temperament might be focused on the rewarding aspect of weight loss through caloric restriction; the attention of those of avoidance temperament may be centered on the perceived threat of weight gain through caloric indulgence.

Another possible moderator in the relationship between disordered eating and implicit caloric-related attitude is reward-based eating drive. This construct is narrower in scope than reward responsiveness because it pertains solely to reward-related eating behaviour. More specifically, reward-based eating drive is an absence of control over eating, lack of satiation, and a preoccupation with food (Epel et al., 2014). Together, these factors comprise a heightened responsiveness to food-related reward. This more specific aspect of reward responsiveness may explain the inconsistent results concerning implicit caloric-related attitudes among individuals with disordered eating. Further, higher reward-based eating drive could account for a more positive implicit attitude toward palatable foods whose consumption is generally experienced as rewarding and which are also typically high in caloric value.

The present study investigated the relationship between disordered eating and implicit caloric-related attitude. Given the inconsistent findings
concerning the direction of implicit attitude toward the caloric value of food among individuals with disordered eating (Hoefling & Strack, 2008; Maison et al., 2001), it was hypothesized that the nature of this relationship might depend on reward responsiveness, a personality construct that also has been found to be inconsistent in this population (Glashouwer et al., 2014; Harrison et al., 2016). A competing hypothesis was that reward-based eating drive might account for the observed differences in implicit attitude. Delineating the nature of the links between these four constructs was the purpose of this exploratory analysis. Specifically, it was assessed whether a broad dimension of personality, reward responsiveness, could account for differences in implicit caloric-related attitude among those high in disordered eating or whether observed differences were simply due to individual variation in responsiveness to food-related reward.

Method

Participants

Undergraduate student volunteers ($N = 100$, 77 women) participated in this study and received bonus points toward their final grades in eligible psychology courses. This study was approved by the Research Ethics Board at Lakehead University, and participants provided informed consent prior to their participation. Participants ranged from age 17 to 44 ($M = 20.12$, $SD = 3.66$). Regarding ethnicity, 75% of the sample identified as European Canadian/White, 7% identified as Asian, 6% identified as African Canadian/Black, and 12% identified other ethnicities including Aboriginal/First Nation, Middle Eastern, and Hispanic. In the present study, 12% of participants could be categorized according to scores obtained on the Eating Attitudes Test-26 as demonstrating overconcern regarding dieting, weight, and eating behaviours (Garner, Olmsted, Bohr, & Garfinkel, 1982). As part of a larger study, participants were required to be nonsmokers and not taking antidepressant, hypertension, or cold medication at the time of participation.

Measures

Reward Responsiveness Scale. The Reward Responsiveness Scale (RRS; Van den Berg et al., 2010) is an eight-item self-report measure of reward responsiveness, a personality construct characterized by sensitivity to reward. Half of the items derive from the well-known behavioural approach system reward scale (Carver & White, 1994), with which the RRS is highly associated, $r = .68$. In demonstration of the discriminant validity of this measure, the RRS and the behavioural inhibition system scale by Carver and White (1994) are uncorrelated (Van den Berg, 2010). Respondents indicate the degree to which they agree with each item on a 4-point scale ranging from 1 (strong disagreement) to 4 (strong agreement). Responses are summed to produce a total score where higher values represent greater reward responsiveness. In the present study, the reliability coefficient for the RRS was in the acceptable range (Cronbach’s $\alpha = .84$).

Reward-Based Eating Drive Scale. The Reward-Based Eating Drive Scale (REDS; Epel et al., 2014) is a nine-item self-report measure of reward-based eating drive. The scale is derived from two existing psychometric instruments: the Three Factor Eating Questionnaire (Stunkard & Messick, 1985), and the Binge Eating Scale (Gormally, Black, Daston, & Rardin, 1982), and also includes novel items. Respondents indicate the degree to which they agree with each item on a 5-point scale ranging from 0 (not at all) to 4 (very much). An item average is derived with higher scores representing greater reward-based eating drive. In the present study, the reliability coefficient for the REDS was in the acceptable range (Cronbach’s $\alpha = .89$).

Eating Attitudes Test-26. The Eating Attitudes Test-26 (EAT-26; Garner et al., 1982) is a 26-item self-report measure of disordered eating that consists of three subscales: dieting, bulimia and food preoccupation, and oral control. Respondents indicate the frequency with which they engage in the behaviours described on a 6-point scale where responses “always,” “usually,” and “often” are scored as 3, 2, and 1, respectively, with the remainder of responses scored as 0. Item 26 is reverse scored. Responses are summed to produce a total score where higher values represent greater disordered eating. The EAT-26 has frequently been used in research to measure the construct of disordered eating. In the present sample, the EAT-26 correlated significantly with the Eating Disorder Examination–Questionnaire (Fairburn & Beglin, 2008), $r = .63$. In addition, the reliability coefficient for the EAT-26 was in the acceptable range (Cronbach’s $\alpha = .84$).

Implicit association test. An IAT was completed by participants. The IAT measures the relative strength of association between an attribute dimension and a pair of target concepts (Greenwald, McGhee, & Schwartz, 1998). Concepts in the present study included high- and low-calorie foods. The dimension of interest was participants’ evaluation of these food categories and, thus, the
IAT involved categorizing images of high- and low-calorie foods with positive and negative words. The IAT literature suggests that the categorization of a pair of concepts, along with positive and negative words, can provide inferential data about the relative implicit attitude regarding the concepts on the basis of reaction times and error rates. If the categorization of positive words with low-calorie foods occurred more quickly and with fewer errors than the categorization of high-calorie foods with positive words, it would be assumed that the participant held a relatively more positive implicit attitude toward low-calorie food in comparison with high-calorie food. Images for the IAT were obtained from the Food-pics image database (Blechert, Meule, Busch, & Ohla, 2014). Items were selected on the basis of caloric density. High-calorie items had greater than 200 calories per 100 grams and included a cookie and some cashews, for example. Low-calorie items had less than 50 calories per 100 grams and included a cucumber and some blueberries, for example. A subset of positive and negative words was selected for inclusion from the word lists provided by Greenwald et al. (1998). The positive words used were happy, honest, health, love, peace, cheer, friend, and pleasure. The negative words used included hatred, rotten, filth, poison, sickness, evil, death, and grief. The IAT was administered on a television screen via Inquisit v4 computer software (www.millisecond.com). The participant categorized the stimuli that were presented in the center of the screen using the “E” and “I” keys of a standard computer keyboard.

Procedure

After institutional review board approval (1465300) was given, participants completed an online questionnaire battery including the RRS, REDS, and EAT-26 prior to attending a laboratory session on a separate occasion. During the laboratory session, participants first engaged in a task that involved viewing images of food stimuli, rank ordering the food stimuli by preference, and sampling the foods, as part of a separate study. Participants also engaged in an IAT that was of interest to the present study, which involved the categorization of images by preference, and sampling the foods, as part of a separate study. Participants were exposed to the same food stimuli and samples, which included images and foods that were categorized the stimuli that were presented in the center of the screen using the “E” and “I” keys of a standard computer keyboard.

Results

Descriptive statistics of the psychometrics are presented in Table 1. The descriptive statistics of the RRS reported by Van den Berg et al. (2010) included a mean of 26.1 and a standard deviation of 3.2. These values align with the descriptive statistics obtained in the present study, which are presented in Table 1. The descriptive statistics of the REDS reported by Epel et al. (2014) included a mean of 1.88 and a standard deviation of 0.71, which are similar to the descriptive statistics obtained in the present study (see Table 1). The descriptive statistics of the EAT-26 reported by Garner et al. (1982) for a female university student sample included a mean of 9.9 and a standard deviation of 9.2. These values are similar to the descriptive statistics obtained in the present study, which are presented in Table 1. A significant and positive skew was observed for disordered eating using Z skewness whereby a Z score of +/−1.96 is considered indicative of a significant degree of skew (Field, 2013). Disordered eating, measured using the EAT-26, was subjected to the natural logarithmic transformation, producing a Z skewness statistic of -0.34, and thus remedying the skew. The transformed variable EAT-26ln was retained for subsequent analyses to represent the construct of disordered eating. A significant and positive skew was also observed for IAT. Two was added to the IAT variable to remove negative values. IAT+2 was then subjected to the natural logarithmic transformation.
transformation, producing a Zskewness statistic of -4.65. This log transformation rendered the skew more severe and therefore, the original IAT variable was retained for subsequent analyses.

The possible range of the IAT variable was between -2 and 2, with positive scores indicating a positive implicit attitude toward low-calorie food and negative scores indicating a positive implicit attitude toward high-calorie food. Ninety-four percent of participants obtained a positive IAT score, indicating a positive implicit attitude toward low-calorie food relative to high-calorie food.

Two moderated multiple-regression analyses were conducted using the SPSS PROCESS macro for Model 1 (Hayes, 2013). The first tested whether the regression of IAT score (Y) on disordered eating (X) was moderated by reward-based eating drive (M). The interaction was not significant (Table 2). The second investigated whether the regression of IAT score (Y) on disordered eating (X) was moderated by reward responsiveness (M). The significant interaction between reward responsiveness (RRS) and disordered eating (EAT-26ln) is displayed in Table 2. The relationship between disordered eating and implicit attitude toward the caloric value of food is contingent upon level of reward responsiveness.

Decomposing the interaction into simple slopes revealed that individuals low (-1 SD) in reward responsiveness evidenced a negative association between disordered eating and positive implicit attitude toward low-calorie food, b = -0.15 [SE b = .07], t = -2.26, p = .026; among those high (+1 SD) in reward responsiveness, the association was positive, b = 0.14 [SE b = .05], t = 2.70, p = .008 (see Figure 1).

**Discussion**

The majority of participants demonstrated a positive implicit attitude toward low-calorie food. The strength of the positive implicit attitude toward low-calorie food was most variable among those high in disordered eating. This variance can be explained by individual differences in reward responsiveness. Among those high in reward responsiveness, disordered eating predicted greater strength of positive implicit attitude toward low-calorie food. Among those low in reward responsiveness, disordered eating predicted lesser strength of positive implicit attitude toward low-calorie food. This finding can be understood with reference to the two personality temperaments described by Elliot and Thrash (2002) concerning reward responsiveness and approach-avoidance motivation. Individuals of approach temperament are high in reward responsiveness; they scan their environment for cues of potential reward and engage in reward-seeking behaviours (Elliot & Thrash, 2002, 2010). Applied to disordered eating, where weight loss and caloric restriction might be

### Table 1

**Descriptive Statistics and Intercorrelations**

<table>
<thead>
<tr>
<th>Variables</th>
<th>M</th>
<th>SD</th>
<th>Cronbach’s α</th>
<th>Zskewness</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 RRS</td>
<td>25.36</td>
<td>3.50</td>
<td>.84</td>
<td>-1.89</td>
<td>.15</td>
<td>.09</td>
<td>.16</td>
</tr>
<tr>
<td>2 REDS</td>
<td>1.49</td>
<td>0.85</td>
<td>.89</td>
<td>0.98</td>
<td>.41</td>
<td>-.08</td>
<td></td>
</tr>
<tr>
<td>3 EAT-26</td>
<td>8.39</td>
<td>7.93</td>
<td>.84</td>
<td>6.52</td>
<td>.02</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4 IAT</td>
<td>0.62</td>
<td>0.35</td>
<td></td>
<td>-2.57</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. N = 100. RRS = Reward Responsiveness Scale; REDS = Reward-Based Eating Drive Scale; EAT-26 = Eating Attitudes Test-26; IAT = Implicit Association Test. *p < .05.

### Table 2

**Unstandardized Regression Coefficients for Predicting IAT Implicit Attitude Toward Low-Calorie Food From EAT-26ln as Moderated by REDS and RRS**

<table>
<thead>
<tr>
<th>Variables</th>
<th>b (95% CI)</th>
<th>SE b</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>0.72 (0.41, 1.02)</td>
<td>0.15</td>
<td>4.67</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>REDS</td>
<td>-0.08 [-0.33, 0.17]</td>
<td>0.13</td>
<td>-0.66</td>
<td>.513</td>
</tr>
<tr>
<td>EAT-26ln</td>
<td>-0.02 [-0.21, 0.17]</td>
<td>0.10</td>
<td>-0.22</td>
<td>.823</td>
</tr>
<tr>
<td>REDS x EAT-26ln</td>
<td>0.02 [-0.10, 0.14]</td>
<td>0.06</td>
<td>0.36</td>
<td>.720</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Variables</th>
<th>b (95% CI)</th>
<th>SE b</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>2.25 (1.17, 3.34)</td>
<td>0.35</td>
<td>4.12</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>RRS</td>
<td>-0.06 [-0.11, -0.02]</td>
<td>0.02</td>
<td>-3.02</td>
<td>.003</td>
</tr>
<tr>
<td>EAT-26ln</td>
<td>-1.08 [-1.72, -0.43]</td>
<td>0.32</td>
<td>-3.32</td>
<td>.001</td>
</tr>
<tr>
<td>RRS x EAT-26ln</td>
<td>0.04 [0.02, 0.07]</td>
<td>0.01</td>
<td>3.44</td>
<td>.001</td>
</tr>
</tbody>
</table>

Note. N = 100. REDS = Reward-Based Eating Drive Scale; RRS = Reward Responsiveness Scale, \( R^2 = 0.01; EAT-26 = \) long transformed Eating Attitudes Test-26.

### Figure 1

![Figure 1](https://example.com/figure1.png)

RRS = Reward Responsiveness Scale; EAT-26 = Eating Attitudes Test-26; IAT = Implicit Association Test.
experienced as rewarding, this may manifest as a propensity for low-calorie food in terms of perceptions and behaviours. Previous research has demonstrated the predictive validity of the IAT in terms of actual behaviour in instances where ambivalence may exist regarding food-related attitudes. For example, De Houwer and De Bruycker (2007) found that those adhering to a vegetarian diet held more positive implicit attitudes toward vegetables relative to animal products in comparison with meat eaters. This result has since been replicated by Barnes-Holmes, Murtagh, Barnes-Holmes, and Stewart (2010).

Given the obtained result demonstrating the positive bias to low-calorie foods among individuals high in disordered eating when reward responsiveness is also high, a case can be made for divergent eating behaviour according to reward responsiveness. Specifically, it may be that individuals high in both disordered eating and reward responsiveness adopt a dieting strategy characterized by a motivation to approach low-calorie foods, thus accounting for the strong implicit bias in favour of such foods. A recent study demonstrated the superior performance of implicit attitude toward the caloric value of food in the prediction of participants’ intention to buy the target foods, as compared to the predictive ability of explicit measures (Songa & Russo, 2018). This provides strength to the argument that food-related implicit attitudes can be predictive of eating behaviour and perhaps also of dieting strategy.

Individuals of avoidance temperament, by contrast, are low in reward responsiveness; they scan their environment for cues of potential threat and behave in a way that serves to minimize their chances of encountering such threat (Elliot & Thrash, 2002, 2010). Rather than the hypothesized strategy of approaching low-calorie foods adopted by individuals with disordered eating of approach temperament, individuals of avoidance temperament could, in theory, adopt a dieting strategy characterized by a motivation to avoid high-calorie foods. The direction of implicit attitude would remain positive toward low-calorie food because these two subgroups share the goal of weight loss, which is perceived as attainable through caloric restriction. The strength of the positive implicit attitude, however, would vary due to the differing loci of focus and alternative behavioural strategies. The present findings are a preliminary indicator of this possibility. Further research assessing the relation of the approach and avoidance temperaments proposed by Elliot and Thrash (2010) in the context of disordered eating and implicit bias is required, with the inclusion of an assessment of the specific behavioural approach that is utilized in the pursuit of weight loss.

No moderating effect of reward-based eating drive was observed. This finding supports the hypothesis that the more general personality dimension of reward responsiveness, indicative of approach and avoidance temperaments, accounts for the variance in implicit caloric-related attitude results as opposed to an eating-specific reward sensitivity.

Limitations
The nonclinical sample was obtained from a university student population and, thus, generalizability to more severe eating disorder presentations is unknown. A minority of participants obtained scores on the measure of disordered eating (EAT-26) that suggest overconcern with dieting and weight (Garner et al., 1982). Given the widespread prevalence of subthreshold patterns of disordered eating beginning prior to the age at which post-secondary education is pursued (Chamay-Weber, Narring, & Michaud, 2005), these participants were retained to foster the generalizability of the obtained findings. This measure, however, cannot be used to reliably differentiate those with or without clinical levels of disordered eating. As such, a limitation of the current study is the lack of information gathered regarding the current or historical presence of clinically severe levels of eating pathology. Furthermore, because a convenience sample of university students was recruited, the representativeness of these findings with respect to gender is limited. Although disordered eating is observed with greater frequency in female populations, this issue is also apparent with men and should not be neglected in future research. This research was conducted at a university located in a northern region of Ontario, Canada, and as a result, the ethnic diversity of the participant sample is not representative of what would be observed in more metropolitan areas of Canada or the United States. The generalizability of the findings is therefore limited to a university population in this geographic region and/or a university population with a similar demographic of students.

An index of sensitivity to threat or punishment was not included in the online questionnaire battery, which is a limitation of the present study. The inclusion of this type of measure would have
enabled a more robust assessment of approach and avoidance temperament.

Participants engaged in a number of food-related tasks prior to completion of the IAT. All participants were exposed to the same food images and samples, which minimizes the likelihood of the observed moderation being attributable to an artefact created by participation in food-related tasks prior to their engagement in the task of interest for the present study. In addition, all earlier tasks contained food images and samples comprising a spectrum of foods in terms of caloric value, thus reducing the risk of introducing bias in favor of either high- or low-calorie foods. Nonetheless, the potential influence of participation in the earlier tasks upon the obtained results cannot be isolated nor ruled out.

Finally, it should be noted that there are a number of criticisms of the IAT for the assessment of implicit attitudes. These include questions regarding the stability of the implicit constructs of interest and the potential confounding effect of general processing speed (Blanton & Jaccard, 2008).

Conclusions

Reward responsiveness appears to predict the strength of positive implicit attitudes toward low-calorie food in individuals with disordered eating. Those high in disordered eating and high in reward responsiveness demonstrated the strongest positive implicit attitude toward low-calorie food. By contrast, the positive implicit attitude toward low-calorie food demonstrated by those high in disordered eating and low in reward responsiveness was of the lowest strength. These findings suggest that personality variables affect the behavioural strategy adopted by individuals with disordered eating in the pursuit of their shared goal of weight loss. These differing strategies may have implications for treatment in terms of the potential impact on clinical presentation and targets of treatment. The clinical significance of these results lies in the potentially different manifestations of eating pathology among individuals with clinically severe presentations. Future research should aim to replicate the current findings in a clinical population with attention to differences in symptoms, diagnoses, and response to treatment. For example, it may be that those of high reward responsiveness tend to manifest with more restrictive symptoms of eating pathology, and those with lower reward responsiveness may be more prone to binge/purge symptoms because of their potential orientation toward the target of avoidance.

The findings of the current study suggest that implicit bias toward the caloric value of food among individuals with disordered eating tendencies may vary according to individual differences of the personality variable reward responsiveness. Although the possibility of these individual differences influencing idiosyncratic presentations of eating pathology cannot be ruled out, this was not something that was investigated in the current study. Furthermore, the measurement of implicit bias is not routine practice in psychological assessment. Future research is recommended to assess whether the individual differences observed in the current study (a) are related to individual differences in disordered eating presentation; and (b) whether these hypothetically different presentations necessitate individually tailored intervention. Nonetheless, the current results show that a personality variable can influence a factor—implicit bias—that has previously been shown to predict eating behaviour. Given this finding, it is recommended to individuals who are working with college or university students—some who may be experiencing symptoms of disordered eating—to be sensitive to the potentially wide array of individual differences that can influence beliefs and attitudes toward food. The current findings highlight the importance of conducting a thorough psychological assessment and taking individual characteristics into account while formulating a plan for intervention. The consideration of individual characteristics, culture, and preferences is one of the three pillars of evidence-based practice in psychology (American Psychological Association, 2006).

In terms of practical significance to research, these findings may help to clarify the inconsistent results regarding the directionality of implicit attitude to the caloric value of food. Mixed findings have plagued this field of literature. Future research should ensure the inclusion of a measure of reward responsiveness in order to tease apart the effect of the variable of interest from the potential confound introduced by variability in reward sensitivity.

References


Implicit Attitude in Disordered Eating


Author Note. Brittany A. Mascioli, Department of Psychology, Lakehead University; Ron Davis, Department of Psychology, Lakehead University. Special thanks to Psi Chi Journal reviewers for their support.
The concept of self-fulfilling prophecies was first described by Merton in 1948 when he said “...a self-fulfilling prophecy is, in the beginning, a false definition of the situation evoking a new behavior which makes the originally false conception come true” (p. 195). For example, a teacher may believe that a student is not intelligent, leading the teacher to spend less time helping the student succeed academically. Ultimately, this behavior may fulfill the teacher’s expectation of poor student performance. Self-fulfilling prophecies occur all around us. More specifically, they can influence a child’s success in school, dictate the kinds of relationships a person forms, and have

**Doing a 180: Examining the Stability and Reversal of Behavioral Confirmation Effects**

Jennifer L. Mezzapelle and Michael R. Andreychik
Fairfield University

**ABSTRACT.** For decades, social psychologists have examined the concept of behavioral expectancies, also known as self-fulfilling prophecies, and the long-lasting impact that they can have on individuals’ lives. The reversal of such expectancy effects has received much less attention. The present study focused on the questions of how stable behavioral tendencies elicited via self-fulfilling prophecies are, and the ease, or difficulty, with which expectancy-congruent behavioral tendencies can be reversed. To examine these questions, participants completed varying numbers of computerized reaction time tasks against computer opponents. Participants first played 1, 3, or 5 games against opponents who treated them with hostility, followed by a single game against an opponent who treated them with kindness, and finally played against an opponent who displayed neutral behavior toward them. We predicted that the more times participants played against a hostile opponent, the more difficult it would be for the participants to reduce expectancy-congruent hostile behavior through interaction with an opponent who treated them with kindness. Contrary to our prediction, participants who played the most games against hostile opponents before being exposed to a kind opponent were significantly kinder to the final neutral opponent than were other participants who received less hostile treatment ($p = .002$). These results suggest that it may be possible to counteract the negative effects of behavioral expectancies in some cases. Discussion centers on examining connections between this work and scholarship on empathy and altruism, as well as a consideration of future directions suggested by the results.
the potential to lead to the perpetuation and generalization of stereotypes (Downey, Freitas, Michaelis, & Khouri, 1998; Snyder & Klein, 2005). Although it is now well-documented that self-fulfilling prophecies can have powerful effects on the behaviors and life outcomes of others, research into the specifics of their operation is still lacking. In particular, very little work has been done examining the conditions under which the effects of self-fulfilling prophecies can be reduced or even reversed.

In one of the more well-known demonstrations of self-fulfilling prophecies, researchers Rosenthal and Jacobson (1968) conducted a study in an elementary school in which students were administered an intelligence test at the end of the school year. At the start of the following school year, the students’ new teachers were given bogus results about the students’ performances on the intelligence test. The results listed the names of various students who were randomly labeled as “spurters,” which meant that they would make significant academic gains in the coming school year. It was assumed that the students left off the list would not make significant academic gains and would instead progress at a typical pace in the coming year. Subsequent tests conducted at the end of the school year showed that the spurters did, in fact, make academic gains.

Given that the children labeled as spurters were randomly selected, the authors concluded that it was the teachers’ expectations and treatment of the students (i.e., being more attentive to spurters, asking them more questions, and providing them with more helpful feedback) and not any superior natural abilities of the children that led to the academic gains (Rosenthal & Jacobson, 1968). Importantly, however, the teachers were not aware of the impact that their expectations and treatment of the students had on the students’ academic successes. In short, the teachers’ expectations were fulfilled, not because the students were spurters, but because the teachers expected them to be, thus illustrating the impact that self-fulfilling prophecies can have on daily life.

Of course, because Rosenthal and Jacobson (1968) did not follow the students in this study over time, their results cannot reveal whether teachers’ behavioral expectancies had any long-term impact on students. However, additional research has demonstrated that behavioral expectancies can “carry over,” or lead targets to behave in expectancy-consistent ways in contexts in which they did not receive expectancy-congruent treatment (Snyder & Klein, 2005; see also Chen & Bargh, 1997; Snyder, Tanke, & Berscheid, 1977), and in some cases even caused targets to change their self-concept (Kelley & Stahelski, 1970; Snyder & Klein, 2005). For example, treating people as though they are competitive in one situation may in some instances cause broader changes in their self-concept, leading them to see themselves as generally competitive. The expectations that people hold for other people in a single setting (or even a single encounter) can thus extend far beyond that specific setting, potentially transforming them into a certain type of person (e.g., a competitive person, an intelligent person) even if they were not such a person initially. This work also relates to ongoing debates in psychology regarding the nature of personality and its potential for change. Although a full discussion of this (complex) issue is beyond the scope of the present article, the work just reviewed does suggest that, at least in some cases, social interactions can shape or alter aspects of an individual’s personality (Allemand & Martin, 2016; Caspi, Roberts, & Shiner, 2005).

A particularly powerful example of the potentially long-lasting impacts that self-fulfilling prophecies can have on individuals was provided by Snyder and Swann (1978). In Snyder and Swann’s study, participants took part in a reaction time game, the object of which was to outperform their opponent in terms of both speed of reaction and strategic use of a “noise weapon.” The task was done in pairs, with one participant assigned to the role of perceiver and another participant assigned to the role of target. Participants were instructed that, during each trial, either they or their partners would have the opportunity to use a “noise weapon” of six varying intensities to distract the other person. Perceivers who were given the expectation that their partner was hostile prior to beginning the game behaved in a hostile manner toward their partners (i.e., using high levels of the noise weapon), thus leading their target partners to reciprocate and behaviorally confirm the perceivers’ expectations. When the targets who were treated as hostile by their first partner played the same game against a new partner who had no expectation of them, the targets continued to play in a hostile manner. These findings suggest that the initial partners’ expectancies became a relatively enduring part of the targets’ approach to the game.

Taken together, the work of researchers such as Rosenthal and Jacobson (1968) and Snyder
and Swann (1978) has provided some evidence regarding the operation of self-fulfilling prophecies and the powerful influence that behavioral expectancies can have on an individual’s life. However, much less work exists regarding the question of whether there are conditions under which behavioral expectancy effects can be reduced or even reversed. For example, what happens when the spurters discussed above leave their current classrooms? Will they continue to make significant academic gains in a new classroom, away from the expectations of their original teachers? Perhaps even more interestingly, what will happen to the spurters if they are later exposed to teachers who hold different expectations for them? Will the behaviors elicited by the favorable expectations of their initial teachers persist, or will their behavior change to align with the (different) expectations of their new teachers?

Smith, Jussim, and Eccles (1999) provided some evidence relevant to this issue in a long-term study of student performance. The researchers followed groups of students from their sixth and seventh grade years through their senior years in high school. Their results showed that teachers’ expectations for students’ levels of success in mathematics measured in the first year of the study were associated with their performance in mathematics that year, similar to the results observed in Rosenthal and Jacobson (1968). In subsequent years, however, Smith et al. (1999) observed that, whereas the performance of some students could still be predicted by the expectations of their first-year teacher, the performance of other students was no longer related to their first-year teacher’s expectations. These results suggest that, although expectancies can have consistent effects on the lives of the targets, in some cases the effects of others’ expectancies can be undone, at least to some degree, by future interactions with others who do not hold such expectancies (or, perhaps, hold different expectancies).

Although Smith et al. (1999) suggested that the effects of self-fulfilling prophecies can be reversed, their correlational study did not show why or how these expectancy effects can be changed. Additionally, the results only provided information about the permanence and reversibility of behavioral expectancies in an educational setting. As mentioned, self-fulfilling prophecies can occur in a myriad of different contexts, not just education, which warrants an investigation of whether they can be reversed in these other contexts as well. To examine these issues, the current research focused on the questions of at what point expectancy effects become a permanent part of an individual’s behavior within a given setting and at what point expectancy effects can be reversed or eliminated. We took an experimental approach which allowed us to carefully control the type of treatment participants received and thus examined what type of treatment might cause a previously established expectancy effect to be undone. In allowing us to better pinpoint the cause(s) of any reversals in expectancy effects, this experimental approach thus complements the work of Smith et al. (1999), which examined the long-term functioning of these processes in a natural and nonexperimental setting.

The current study was largely based on the previously described work of Snyder and Swann (1978). In our study, participants also played a game with a noise weapon; however, instead of real opponents, they played against preprogrammed computer opponents. To keep responses genuine, participants were made to believe that they were playing against other students in different laboratories. Participants were treated as though they were hostile by a varying number of ostensible opponents (e.g., one, three, or five opponents). Then, they played against a new opponent who served the purpose of trying to reverse the self-fulfilling prophecy by treating participants in a way that was incongruent with the original treatment they received (i.e., a partner who treated them very kindly after they have been treated very aggressively by other partner or partners). Last, participants played the game against a neutral opponent who used moderate levels of the noise weapon. This last game was meant to test whether the varying amounts of hostile treatment could be overridden by subsequent contrasting (kind) treatment. Presumably, the neutral opponent represented a “blank canvas” on which a participant’s preferred method of playing the game could be expressed. For example, if a participant acted in a hostile manner toward the neutral opponent, it would suggest that the hostile treatment received from the first opponent(s) could not be overridden by the kinder treatment that followed. On the other hand, if a participant behaved kindly toward the neutral opponent, it would suggest that the hostile treatment was reversed by the kind treatment. The hypothesis was that the more hostility-inducing treatment participants received, the less likely it would be that their hostile behavior could be undone by interacting with a kindness-inducing
Behavioral Confirmation Stability and Reversal | Mezzapelle and Andreychik

opponent, and the higher the levels of the noise weapon they would use against their neutral opponent. On the other hand, the less hostility-inducing treatment individuals received, the more likely it would be that their hostile behavior could be undone by the kindness-inducing opponent, and the lower the levels of the noise weapon they would use against their neutral opponent.

Method

Participants

The sample consisted of 105 undergraduate students from Fairfield University who participated for course credit or extra credit in their General Psychology or Statistics courses. The sample was 79% women and 21% men with a mean age of 19 (SD = 1.13, range 18–22). A power analysis conducted using G*Power (Faul, Erdfelder, Buchner, & Lang, 2009) indicated that a sample size of N = 158 was necessary to provide 80% power (Cohen, 1992) to detect a medium-sized effect of the independent variable employing an α of .05. However, because we were limited by the relatively small size of our participant pool, our sample size was N = 93. As such, the study only achieved 55% power to detect an effect of our independent variable. This limitation will be further discussed in the Discussion section.

Materials

Participants completed a personality measure at the beginning of the study to determine their baseline levels of aggression and competitiveness. This personality measure was composed of four competitiveness-relevant items taken from the Big Five Inventory (e.g., “I see myself as someone who starts quarrels with others;” John, Donahue, & Kentle, 1991), along with two additional items that were written specifically for this study that focused on competitiveness and aggression (e.g., “I see myself as someone who is often competitive”). Sixteen additional items were also taken from the Big Five Inventory and used as filler items to distract participants from our true interest in competitiveness. The six items specifically relevant to competitiveness were averaged, after appropriate reverse scoring, to provide an overall measure of participants’ baseline levels of aggression (M = 2.12, SD = 0.58, α = .57). The complete personality assessment can be found in the Appendix.

Although we had initially planned to use baseline levels of aggression as a covariate in our analyses, given the relatively low reliability of the measure, we decided to omit the covariate from the analyses. The pattern of results reported below is the same regardless of whether baseline aggression is included as a covariate.

Following the personality assessment, participants played reaction time games against varying numbers of computerized opponents. Participants looked at a black fixation point on the computer screen and were told that their job was to press the spacebar as quickly as possible when a red dot appeared in place of the fixation point. The computer randomly selected the amount of time between the appearance of the fixation point and the appearance of the red dot.

Each game consisted of 24 trials of reacting to the red dot. The 24 trials were broken up into 8 blocks of 3 trials each. At the beginning of each block, the option to employ the noise weapon alternated. For example, if a participant selected the level of the noise weapon to use against an opponent at the start of the first block, then the computer would select the noise level in the next block. When it was the computer opponent’s turn to employ the noise weapon, the participant would hear the noise as the red dot appeared. This part of the study was based on the method of the experiment conducted by Snyder and Swann (1978), who used a similar methodology to successfully produce self-fulfilling prophecies in the laboratory. The noise weapon was offered in nine different intensities, ranging from 49 decibels to 81 decibels, with each level being 4 decibels higher than the level before it. An intensity of 1 was described as “inoffensive,” 5 was considered “distracting,” and 9 was described as “offensively irritating and annoying” (Snyder & Swann, 1978). Prior to the actual games, participants heard samples of the inoffensive, distracting, and offensively irritating intensities. Although we used the same language to describe the noise intensities, we chose to use nine distinct noise levels, as opposed to the six levels used in Snyder and Swann’s experiment (1978). This allowed us to create more defined category labels for levels of noise selection. Within the games, noise weapon levels 1–3 were considered kind, levels 4–6 were neutral, and levels 7–9 were considered hostile. As such, each opponent that participants encountered was labeled “hostile.”

1 Although 105 participants took part in the study, the data of 12 participants were excluded from analyses, leaving us with the data of 93 participants.

2 Complete Inquisit scripts for the study can be accessed at osf.io/gqk6s.
that they were playing against other students who were set up in different laboratories on the floor. In reality, they were playing against preprogrammed computer opponents. Participants were randomly assigned to one of three conditions. Participants in the first condition played only one game—recall that one game consists of 8 blocks of 3 trials each—against a hostility-inducing opponent. We refer to this condition as the minimal hostile treatment condition \((n = 36)\). In the second condition, the moderate hostile treatment condition \((n = 35)\), participants played three games against hostility-inducing opponents. Participants in the final condition, the maximum hostile treatment condition \((n = 35)\), played five hostility-inducing games. In all three conditions, participants were told that they were playing against a different opponent in each different game. The computer opponent always employed the noise weapon in the first block of the first hostile game, and control of the noise weapon alternated between the participant and opponent in each subsequent block. In all remaining hostile games, either the computer or the participant was randomly selected to employ the noise weapon in the first block, and control of the noise weapon again alternated between participant and opponent in subsequent blocks.

Following these games with the hostile opponent(s), all participants next played one game against a kind opponent. The computer opponent always employed the noise weapon in the first block of this game. This game was meant to counteract the behavioral expectation of the previous games. Participants then played a final game against a neutral opponent. Participants always employed the noise weapon in the first block of this neutral game. This was done so that participants could “set the stage,” so to speak, for the game, and it eliminated the potential problem of participants simply mimicking the choices of their present partners.

Following the completion of the study, participants were debriefed verbally and in writing. During debriefing, some participants reported being suspicious of the cover story of the study. These suspicions were noted by the researchers and data from these participants were excluded in analyses.

**Results**

Of the 105 participants, 12 of them reported during debriefing that, while participating in the study, they were suspicious of the cover story and suspected that they were playing against computer
participants were excluded from the data analyses, leaving 93 participants.\footnote{Additional analyses were done with the full sample of 105, including participants who reported suspicion. Although significance levels changed slightly, the overall pattern of results was the same.} Data were then analyzed using a Between-Subjects Analysis of Variance (ANOVA), which allowed us to see what effect amount of hostile treatment had on selection of noise weapon level.

The results focused on the levels of hostility that participants displayed in the neutral game (see Figure 1). We had predicted that participants in the maximum hostile treatment condition would exhibit the greatest hostility toward the neutral opponent, those in the moderate hostile treatment condition would exhibit a moderate level of hostility, and those in the minimum hostile treatment condition would exhibit the lowest level of hostility. Although amount of hostile treatment did have a significant impact on what level of the noise weapon participants chose to use against their neutral opponent, $F(2,93) = 5.22, p = .007, \eta^2_{partial} = .101$, as can be seen in Figure 1, the nature of this effect was different than what we had predicted.

In particular, follow-up pair-wise comparison tests (see Table 1) showed that participants in the minimum and moderate hostile treatment conditions did not differ significantly in the noise levels they chose to use against their neutral opponents ($p = .167$). Likewise, although the means in the minimum and maximum hostile treatment conditions indicated that participants in the minimum hostile treatment condition used slightly higher levels of the noise weapon against their neutral opponent than participants in the maximum condition did, the difference between the minimum and maximum hostile treatment conditions was not significant ($p = .083$). Finally, the analysis also showed that there was a significant difference between the moderate and maximum hostile treatment conditions, ($p = .002$). The data suggested that participants who were given maximum hostile treatment used significantly lower levels of the noise weapon against the neutral opponent than those in the moderate hostile treatment condition. In summary, the minimum hostile treatment condition did not differ from the moderate or maximum hostile treatment conditions, whereas the maximum hostile treatment condition resulted in significantly less hostility toward the neutral opponent than the moderate hostile treatment condition did.

\begin{table}
\centering
\caption{Mean Differences of Noise Weapon Use Between Conditions}
\begin{tabular}{llll}
\hline
Condition (I) & Condition (J) & Mean Difference (I-J) & Significance (p) \\
\hline
Minimum & Moderate & -0.66 & .167 \\
 & Maximum & 0.84 & .083 \\
Moderate & Minimum & 0.66 & .167 \\
 & Maximum & 1.50$^\star$ & .002$^\star$ \\
Maximum & Minimum & -0.84 & .083 \\
 & Moderate & -1.50$^\star$ & .002$^\star$ \\
\hline
\end{tabular}
\footnotesize{Note. $^\star$ significant at $p < .01$}
\end{table}

\begin{figure}
\centering
\caption{Mean noise weapon levels administered to neutral opponent by each condition. Error bars represent the standard error within each condition.}
\end{figure}

\section*{Discussion}
We had hypothesized that the more hostile treatment participants received from others in a computerized reaction-time game, the harder and less likely it would be to eliminate or reverse the impact that hostile treatment had on the hostility of participants’ own responses toward a neutral opponent. However, the findings suggested that participants who received the maximum level of hostile treatment were the kindest toward the neutral opponent. These results do suggest that in an aggression or hostility-related situation, some expectancy effects can, in fact, be reversed, just as they can be in an educational setting (Smith et al., 1999). Although the results did not support our hypothesis, they provide the opportunity to delve deeper into the nature of self-fulfilling prophecies and examine factors that might have played a role in why we found the results that we did. In this discussion, we review the limitations of this study and offer a few possible explanations for these findings. These explanations, however, are simply starting hypotheses that would require more research to...
A true discussion of our results requires first taking into consideration the limitations of our study. As previously mentioned, because we were limited in how large a sample we could test because of the relatively small size of the undergraduate participant pool from which our participants were drawn, a power analysis showed that our study only achieved 55% power to detect an effect of our independent variable. This limitation highlights the need to conduct this study again in the future with a larger sample, enabling us to obtain more accurate estimates of the effects of our independent variable. Another limitation was that participants’ ages and genders were only collected from the information each participant reported as a member of the participant pool, rather than being collected as part of the study’s data. This means that we were not able to provide gender and age compositions separately for each condition, nor could we analyze any possible effects that these individual-difference variables might have had on the results. However, because the majority of our sample were women, there might not have been enough men to make a gender comparison meaningful in any case. Still, any future version of this study should be sure to collect this information within the study so that the potential effects of these variables can be examined.

It is also worth noting that our measure of participants’ pre-existing levels of competitiveness was composed of a mix of items taken from existing personality instruments and items created specifically by us to relate directly to competitiveness. Of course, modifying existing instruments or creating new questions can affect the psychometric properties of those instruments, so future work should attempt to replicate these results using an existing psychometrically sound measure of dispositional aggression or competitiveness.

Beyond the limitations of our study, there are many possible factors that could have influenced the results. One such factor is empathy. Empathy is the ability to take on the emotions of another person as one’s own (Andreychik & Migliaccio, 2015). For example, empathetic people may find themselves feeling sad when a friend has lost a job. A large part of empathy involves perspective-taking (i.e., trying to imagine what it would feel like to be in the position of the other person). Myers, Laurent, and Hodges (2014) suggested that when faced with another person’s suffering, individuals who are instructed to imagine themselves in the position of the sufferer (imagine-self) are more likely to be empathetic and helpful than individuals who are instructed to instead imagine the feelings of the other person (imagine-other). We never gave any of these instructions to participants, but it is possible that, for participants in the maximum hostility condition who had just been repeatedly treated as though they were hostile, this kind of imagine-self perspective-taking occurred naturally when participants encountered the final neutral opponent. If participants in this condition felt uncomfortable or unpleasant, their recurrent experience of being the target of others’ hostility might have made them more likely to put themselves in the position of their opponent. If this is the case, participants might consider how badly they themselves felt when someone used high levels of the noise weapon against them, resulting in kinder and more empathetic treatment towards the neutral opponent in an attempt to prevent this opponent from experiencing the same discomfort.

Along the same lines, Lim and DeSteno (2016) have shown that individuals who have encountered adversity have higher levels of empathy, which may generate more compassion toward the plights of others. Although the levels of noise that participants heard cannot really be labeled “adversity,” the phenomenon might have operated on a smaller scale in this instance. Perhaps the repeated experience of being the target of hostile treatment led to elevated empathy and motivation to be more compassionate toward their opponent. Of course, these ideas about the effect of empathy on the results are simply hypotheses. A greater understanding of the relationship between empathy and self-fulfilling prophecies would warrant further research. One solution to this might be to replicate the study, this time including measures of empathy and perspective-taking.

Another possible explanation for this study’s findings lies in the concept of reciprocal altruism. Reciprocal altruism is based on the motivation for people to help those who they believe can help them in return in some way. Most existing research on reciprocal altruism focuses on it as a basis for friendships and familial relationships. This is mostly due to the fact that these kinds of close relationships provide ample opportunity for kindness to be repaid (Rotkirch, Lyons, David-Barrett, & Jokela, 2014). However, given the high probability of discomfort in this study, it is possible that reciprocal altruism could manifest itself in a brief interaction such as this one. Here, participants might have believed that, if they were hostile toward the neutral opponent, their...
opponent might reciprocate high levels of the noise weapon. As a result, participants might try to avoid the discomfort that would be caused by yet another game involving hostile levels of the noise weapon by being kind to their opponent.

If the observed pattern of results is the product of reciprocal altruism, only participants in the maximum hostile treatment condition took part in this altruistic behavior. This may be due to the role that rational judgment plays in reciprocal altruism. Korsgaard, Meglino, Lester, and Jeong (2010) have shown that one of the driving factors in reciprocal altruism is the use of rational judgment, which bases decisions on past experiences and expectations for the future. Participants in the maximum hostile treatment condition had an increased exposure to hostile treatment compared to the other conditions, possibly giving them a higher anticipation for future aggression and discomfort. This high anticipation, in turn, may serve as motivation for these participants to try to avoid such hostility by being kind to the final opponent in the hopes that the opponent would reciprocate this kind behavior.

It is also possible that our pattern of results may be specific to the particular characteristic on which we chose to focus. As previously mentioned, self-fulfilling prophecies occur within many contexts, such as education, relationships, and stereotype generalization (Downey et al., 1998; Snyder & Klein, 2005). However, despite the varied contexts in which self-fulfilling prophecies occur, most of the research that examines the lasting stability of self-fulfilling prophecies has been done within the context of education and has focused on the trait of intelligence, specifically mathematical intelligence (Smith et al., 1999). It is possible that these expectancies of intelligence operate differently than expectancies regarding hostility and aggression. Given that our hypothesis was partially based on the literature that used education and mathematical intelligence, this could explain why our results were contrary to the hypothesis and the existing literature. To see if this is the case, a study similar to the one presented here could be conducted with a focus on academic performance. Instead of using the expectation of hostility with a reaction time task, the new study could give the expectation of a particular facet of intelligence, or lack thereof, with some kind of problem-solving task or a test of mathematical ability. We would then be able to examine whether those results are more similar to the results of this study or the one conducted by Smith et al. (1999).

Furthermore, self-fulfilling prophecies that occur in an educational setting, like a classroom, may have specific characteristics that distinguish them from behavioral expectancies taking place in other contexts including a laboratory setting. For example, an expectancy in a classroom might be held and acted on repeatedly by a single person, namely, a teacher. In the present study, most participants encountered a various number of opponents who acted on an expectation on one single occasion. It is possible that having multiple interaction partners versus one repeated interaction partner could influence how strongly an expectation is internalized, which may in turn affect the self-fulfilling prophecy’s reversibility.

A better understanding of the influence of these differences would require further investigation. One option is to conduct a study such as the one suggested above in which intelligence or problem-solving expectancies are examined in a laboratory. Another is to find a real-world context in which aggression-relevant expectancies can be examined such as sports. A future study could focus on a coach’s expectancies for sports players’ aggression levels over a period of time before seeing if the initial coach’s expectancies could be undone by a subsequent coach’s contrary expectations. This would mirror the repeated interaction with a single person exhibited by a teacher in a classroom. Another study could follow similar expectancies held by multiple referees or sporting officials about players’ levels of aggression. This would reflect the effect of expectancies imposed by multiple interactions analogous to those in our study. Looking at these two suggested studies side-by-side could offer new insight into the impact of number of interaction partners within the same context.

Additionally, it is important to consider possible differences in level of vulnerability between the sample that was used in our study and the samples used by studies presented in existing literature. There is evidence to suggest that the more vulnerable people are, particularly related to their self-concept, the more susceptible they are to the effects of self-fulfilling prophecies (Smith et al., 1999). Some potential factors that might make an individual predisposed and vulnerable to self-fulfilling prophecies are age, perceived status of the person(s) holding the expectancies, lack of self-efficacy, or being in a new or unfamiliar environment. Children may be more vulnerable to the effects of self-fulfilling prophecies because they typically do not have a firmly developed self-concept...
(Smith et al., 1999). This means that the children used in the school-based studies referenced in this article might have been more easily impacted by their teachers’ expectations. If children do not have a preestablished idea of themselves as intelligent individuals, they are less likely to resist being treated as though they are unintelligent. On the other hand, the college students who made up our subject pool are more likely to have a more defined self-concept. As a result, they are more likely to behave in a way that is consistent with their existing self-concept, rather than behaviorally confirming an expectation that may be inconsistent with that self-concept (Snyder & Klein, 2005). Essentially, if they strongly believe that they are not hostile and aggressive, they will make the effort to appear kind when faced with the expectation of hostility. Participants in the maximum hostile treatment condition might have been especially motivated to reassert their nonhostile self-concepts because these nonhostile self-concepts were especially threatened by the continued hostile treatment they received from their partners. This issue could be examined by repeating this study with different age groups.

As with any research study, it is of the utmost importance to conduct this kind of research in an ethical manner. As discussed, some self-fulfilling prophecies can have potentially harmful long-term implications. Thus, the development of future studies must take care to protect participants from such implications by executing the research in a way that minimizes the possibility that any negative effects will transfer to life outside of the study. This includes obtaining the informed consent of the participants or participants’ guardians and following the study with a detailed debriefing.

Although the findings of this study were contrary to what we had expected, the results still add to the vast amount of literature that already exists regarding self-fulfilling prophecies. The results demonstrate that there is still a great deal of work that needs to be done to fully understand the nature of these expectancies and their effects. More specifically, combined with the existing literature, these results demonstrate that some self-fulfilling prophecies can be overturned, but the ability to reverse them may be dependent on a variety of factors such as the context in which the expectation occurs, who is acting on it (and how often), how many people hold the expectation, and the trait or behavior on which the expectancy focuses. Furthermore, understanding the different ways in which self-fulfilling prophecies can operate is absolutely crucial to be able to effectively reduce the harmful effects that others’ negative expectations can have in the real world.

References


Author Note. Jennifer L. Mezzapelle, Psychology Department, Fairfield University; Michael R. Andreychik, Psychology Department, Fairfield University.

This research was funded by a Mancini Scholarship Award, awarded to Jennifer L. Mezzapelle. Special thanks to Psi Chi Journal reviewers for their support.

Correspondence concerning this article should be addressed to Jennifer Mezzapelle at jennifer.mezzapelle@student.fairfield.edu or Michael Andreychik at mandreychik@fairfield.edu.

APPENDIX

Here are a number of characteristics that may or may not apply to you. For example, do you agree that you are someone who likes to spend time with others? Please select the number below each statement to indicate the extent to which you agree or disagree with that statement.

<table>
<thead>
<tr>
<th>Statement</th>
<th>Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Disagree strongly</td>
<td>1</td>
</tr>
<tr>
<td>Disagree a little</td>
<td>2</td>
</tr>
<tr>
<td>Neither agree nor disagree</td>
<td>3</td>
</tr>
<tr>
<td>Agree a little</td>
<td>4</td>
</tr>
<tr>
<td>Agree strongly</td>
<td>5</td>
</tr>
</tbody>
</table>

1. Likes to cooperate with others’ (R)
2. Is often competitive’
3. Has an assertive personality’
4. Starts quarrels with others’
5. Would do almost anything to avoid conflict’ (R)
6. Is considerate and kind to almost everyone’ (R)
7. Is talkative
8. Does a thorough job
9. Is depressed, blue
10. Is original, comes up with new ideas
11. Can be somewhat careless
12. Is curious about many different things
13. Is full of energy
14. Is a reliable worker
15. Is ingenious, a deep thinker
16. Tends to be disorganized
17. Tends to be lazy
18. Is inventive
19. Perseveres until the task is finished
20. Does things efficiently
21. Is easily distracted
22. Is sophisticated in art, music, or literature

Note. The 22 items in this personality assessment were arranged into a randomized order and then presented in the same randomized order to all participants. Items marked with a * were used to compute the competitiveness index. Items marked with a + were created by the researchers and did not come from the Big Five Inventory. Items marked with (R) were reverse-coded.
Reading is a crucial task that is accomplished with relative ease on a daily basis. Once an individual develops into a proficient reader, this becomes an automatic process. This automaticity has been clearly illustrated by results of the classic Stroop Task, which demonstrated that people are unable to see words without decoding them (Stroop, 1935). A widely shared Internet meme referencing a nonexistent study suggests that this automaticity equally applies to both correctly ordered and scrambled text. The meme claims: “Aoccdrnig to a rscheearch at Cmabrigde Unervtisy, it deosn’t mttar in waht oredr the ltteers in a wrod are”. This is, however, a hoax. Although stimulus and context effects on reading have been explored, no Cambridge University study has shown that letter position in the words people read is irrelevant for word recognition (Davis, 2003).

The closest findings, and possibly the misinterpreted basis for this claim about scrambled word identification, appear to come from one study. Rawlinson (1976) found that, when participants read passages of text where both the first two and final two letters of words were fixed but the middle letters were randomized, there was very little impact on reading comprehension. In fact, not all participants noticed that many of the words contained scrambled letters. But comprehensibility of the sentence as a whole does not automatically equate to easy comprehension of each individual word. Nor does it assume that speed of processing is unaffected, as is suggested by the aforementioned meme.

Although this “Cambridge study” is not real,
attempts have been made to explore the factors that influence people’s ability to read manipulated text. Much of this research has focused on the effects of either replacing or removing individual letters in words (e.g., Grainger, Granier, Farioli, Van Asche, & van Heuven, 2006; Rayner & Kaiser, 1975) or of transposing letters (e.g., Christianson, Johnson, & Rayner, 2005; Perea & Lupker, 2003) rather than truly scrambling the text. These studies generally had the larger objective of validating a range of models of word identification. The goal of the present research was to examine the effect of both alteration type and prior sentence context on scrambled word identification.

The Influence of Text Alteration on Word Processing
In addition to people’s ability to read words that have been altered (Rawlinson, 1976), a range of priming studies have demonstrated that individuals are influenced by exposure to words with transpositions (created by flipping letter positions) even if they are not explicitly aware of their presentation. Perea and Lupker (2004) asked participants to complete a lexical decision task following a prime. They presented participants with a lowercase prime for 50 milliseconds (that the readers were unaware of seeing) followed by the target word in all capitals. The primes were the exact target (e.g., candle for the target CANDLE), had one letter replaced (e.g., candge), had a transposition at positions three and five (e.g., caldne), or had a double substitution in those positions (e.g., candge). Reaction times on the lexical decision task were faster for the transposition prime than for the double substitution prime but only when consonants (but not vowels) were manipulated. This suggests that words with transpositions activate their base word to a greater degree than do two-letter different nonwords. This same effect can be seen for semantically related primes. When priming the semantically related word light, words with an internal transposition (e.g., heavy for heavy) were more effective primes than words with replacements (e.g., heavy, Perea & Lupker, 2003).

Edited primes also influence reaction times when some of the letters are missing entirely. Grainger et al. (2006) primed target words in a lexical decision task with edited versions of the target. Instead of transposing letters, individual letters were replaced with dashes or were completely removed. For example, the word apricot was preceded by the prime a-ri-c or ariet. Both primes led to shorter reaction times than a completed unrelated prime. This was even the case when the dashes were in the wrong location such as in ar-i-cet. The ability of briefly presented words with alterations to prime a target word demonstrates that a word does not have to be presented in its original, fully accurate form in order for the reader to be influenced by it.

When alterations are made to written text, the specific location of the change is critical for word identification. For example, transpositions occurring across morphemes are less beneficial primes than those where the transpositions occur within a morpheme (e.g., sunshine vs. sunhsine, Christianson et al., 2005). A prime created by crossing the boundary between the two halves of a compound word is no more beneficial than a prime created by substituting letters (e.g., subshine). This effect, however, is not always replicated (Rueckl & Rimzhim, 2011; Sánchez-Gutiérrez & Rastle, 2013). Duñabeitia, Perea, and Carreiras (2014) argued that the discrepancy in the literature might be explained by differences in reading speed. They found that cross morpheme transpositions slowed reaction times for faster than average readers, but that there was no difference for slower than average readers. They argued that faster readers may use a slightly different strategy for word recognition than do slower readers.

In the context of straight reading tasks, rather than priming tasks, the beginning of a word has been shown to be particularly important. Rayner, White, Johnson, & Liversedge (2006) recorded reading times for sentences where the content words had letter transpositions such that the order of two adjacent letters were reversed. They found that transpositions that occur at the beginning of a word (e.g., oslwre for solve) were more problematic than those that occur in the middle of a word (e.g., slowe or end of a word (e.g., soler). Although people may be able to read text with transposed letters, this does not occur without a cost to speed of processing (as was erroneously suggested by the fake “Cambridge Study”). A similar outcome occurs when letters are entirely replaced (Rayner & Kaiser, 1975). These findings suggest that the beginning of a word provides more critical information for word identification than does the middle or the end of the word. This is supported by the finding that the initial letters of a word are the easiest to recognize after short presentation durations (Adelman, Marquis, & Sabatos-DeVito, 2010).

Although the beginning of a word is important for identification, some evidence has also suggested that the end of a word may hold a privileged
position in word identification. In a semantically related priming task, Perea \& Lupker (2003) found that, although internal transpositions (e.g., heavy \textit{for} heavy) primed a target word, end transpositions (e.g., heavy) did not. In this context, words with end position transpositions failed to activate semantically related words. Evidence for the relative importance of both the beginning and end of a word comes from a letter identification task. McCusker, Gough, and Bias (1981) presented participants with four-letter words and were cued to name a single letter in the word. The two internal or two external letters appeared 50 milliseconds in advance of the rest of the word or all letters appeared simultaneously. When the whole word was presented at once, response times to name individual letters were faster for the external than for the internal letters, suggesting that external letters are easier to detect. Additionally, participants showed greater overall facilitation for the task when the outside letters appeared first than when the internal letters appeared first.

One of the main goals of the word recognition literature is to test the predictions of a range of computational models for word recognition. These mathematical models rely on input from a lexicon (or mental vocabulary) to learn through experience how to recognize words. The models are designed to simulate what is actually happening in the brain when we are exposed to a written word. A strong model, therefore, must be able to account for the fact that a reader may be able to recognize a word even if it is partially occluded (such as by a coffee stain), when an individual letter is missing (such as a spelling error), or even when letters are reordered. For this reason, a simple model that has strict rules about the location of letters in a word is not sufficient. For example, if the letter \textit{b} is not in the second position of \textit{above}, such as in the transposed example \textit{abobe}, then a strict model would never recognize it.

To account for the fact that words with alterations or deletions can activate their primes (Perea \& Lupker, 2004), modern computational models have taken a more flexible approach to letter position. Two such models are the \textit{sequential encoding regulated by inputs to oscillations within letter units model} (SERIOL; Whitney, 2001) and the \textit{spatial coding model} (Davis, 2010). Both approaches allow for flexibility in letter position. The SERIOL model, for example, does not expect letters to be in a specific position; rather, it recognizes those individual letters based on activation of specific features (i.e., shapes) in each letter (Whitney, 2001). According to the spatial coding model, instead of strictly requiring correct letter position, the location of each letter in a word is seen as having a degree of uncertainty. At the same time, this approach gives priority to the positioning of external letters in a word (Davis, 2010). This allows the model to predict that words where external positions are held constant (i.e., alternations happen within the word rather than at an end) will be seen as more similar to the base word (and thus provide better priming) than those where the external letters are not maintained. Generally, the finding that modifications altering the ordering of letters are more effective primes than those which change the identity of some letters supports models of word identification that do not require specific letter position information and instead allow for some position independence (Perea \& Lupker, 2003).

**The Role of Syntactic and Semantic Context in Word Processing**

Although the studies mentioned previously have generally focused on priming tasks, words are rarely encountered singularly and instead are read, or heard, within the context of a sentence. All well-formed sentences meet a set of linguistic rules. For example, the sentence, “The cat chased the white rat,” is syntactically acceptable because it follows the rules for the appropriate arrangement of words. It also makes sense semantically because the individual words come together to make a meaningful whole.

The syntactic and semantic contexts that the preceding words in a sentence create can influence the perception of an individual word by building up a set of expectations about what is to come next. This expectation effect has been shown to facilitate spoken word recognition (Miller, Heise, \& Lichten, 1951; Miller \& Isard, 1963). There are also effects of different levels of acceptability in reading. For example, sentences that conform to “canonical” word order of a language elicit faster responses than those that employ an acceptable, but less common, word order (Tanaka, Tamaoka, \& Sakai, 2007).

One way that a sentence’s syntactic context influences word identification is that the sentence stem determines what types of words are allowable. For example, after the stem “The girl drank the,” the word \textit{lemonade} is both syntactically and semantically appropriate, but the word \textit{sleeping} is not possible. The frame provides a situation in which a noun is expected (or possibly an adjective before
a noun). Wright and Garret (1984) examined how these expectations influence speed of word identification. Participants saw sentence fragments with a final word as the target for a lexical decision task. The final word was either a verb or a plural noun and either did or did not fit into the preceding context. They found that reaction times were faster for the syntactically acceptable endings (e.g., “The man spoke but could not COMPETE” or “Just at the time of ENTRIES”) than for the syntactically illegal endings (e.g., “The man spoke but could not ENTRIES” or “Just at the time of COMPETE”). The same results have been found in a word naming task (West & Stanovich, 1986). This demonstrates that when people read an individual word in a sentence, they are influenced by the expectations from the syntactic context such that it is easier to process words that would be possible in that situation.

Although a range of words may be syntactically possible at the end of a sentence stem, some options are privileged. For example, although “The boy enjoyed eating the earthworm” is possible and acceptable, it is much less expected than “The boy enjoyed eating the chocolate.” In this way, the semantic information in the situation can help to constrain the possible upcoming words. Schwanenflugel and Shoben (1985) presented participants with sentences that concluded with either a highly expected word or an unexpected, but semantically related, word. Words that fit the expectation were processed more quickly than the less expected (but just as acceptable) words. Evidence from electroencephalography (EEG) studies has also shown that listeners are able to use sentence context to predict upcoming words. An EEG records event-related potential (ERPs), which measures brain activity in response to a stimulus. When listeners are exposed to an unexpected stimulus, they experience a larger N400 (a negatively polarized ERP 400 milliseconds after stimulus onset) than they do for an expected stimulus. Van Berkum, Brown, Zwitserlood, Kooijman, and Hagoort (2005) examined whether an N400 effect could be elicited even before listeners heard an unexpected word. Because nouns in Dutch have a fixed grammatical gender, any associated adjective must have the appropriate gender markings. Van Berkum and colleagues exposed listeners to sentences that strongly predicted an upcoming noun, but which were preceded by an adjective that either did or did not have the appropriate grammatical markings. An N400 effect occurred when hearing an adjective that did not match the gender for the predicted word. This indicates that, by the time the adjective was heard, the brain had already guessed what the upcoming noun would be and was surprised by the incongruent gender of the adjective.

The Present Studies
In the present research, participants were asked to determine the identity of a scrambled word located at the end of a sentence. This design allows for easy manipulation of a range of factors that may influence a reader’s ability to complete the task. Our goal was to explore two factors that may influence both participants’ overall ability to read scrambled words and the speed at which this occurs. Specifically, we focused on the type of word scrambling and the contextual and predictive power of the sentence in which the scrambled word is found. Although these two broader factors have been explored separately, our goal was to examine these factors simultaneously in a reading task to allow for examination of both their individual effects and how they may interact. Although a reader is generally asked to read a correctly formed word, performance on a scrambled word identification task may help to further explain the processes by which people read. The comparison of manipulations within a target word and in the sentence itself that do and do not harm reading speed and accuracy may help identify the most critical factors for word identification.

The effect of scrambling style was examined in both experiments by comparing the ability of participants to recognize scrambled words when either (a) the first and last letters of the word were held in the correct position with the middle scrambled or (b) when all letters were randomized. This is a stricter version of the manipulation used by Rawlinson (1976) in that it provides significantly less information to the reader because only two, not four, of the letters in the word are held constant. Also, unlike Rawlinson’s study, we focused on accuracy and speed rather than overall comprehension. Studies using letter transpositions and letter replacements have shown that the beginning of a word may provide more crucial information for word identification than the middle or the end of the word (Perea & Lupker, 2003; Rayner & Kaiser, 1975; Rayner et al., 2006). Although previous studies have mostly focused on the importance of having both of the first two letters of the word in the correct order, we only held the first letter constant. Our goal was to explore whether the correct position of the first letter alone, in conjunction with proper
placement of the final letter, would significantly impact participants’ ability to decode scrambled text. Based on the demonstrated value of both the beginning and end of a word for identification, it was hypothesized that holding the first and last letter constant should greatly improve both accuracy and speed of decoding as would be predicted by the spatial coding model of word recognition (Davis, 2010).

We also examined the role of expectation and context from the preceding words in the sentence. Although a scrambled word might be examined in isolation, the vast majority of our daily word identification comes in the context of a word in a sentence. Previous research has shown the value of prior context for word identification (e.g., West & Stanovich, 1986; Wright & Garret, 1984) because this information may help readers predict what word will come next. Given these findings, we explored whether this would be true not just for correctly presented words, but also for scrambled word identification. In Experiment 1, sentence context was examined by manipulating the order of the words in the sentence. Although the scrambled word always appeared at the end of a sentence, the words of the sentence itself were sometimes reordered. This manipulation would, therefore, inhibit the ability of the reader to take full advantage of the contextual information in the text for predicting the identity of the scrambled word. If participants wished to make use of the entire sentence, they would have to take the extra time to first unscramble the sentence itself (assuming they even recognized that it made a proper sentence). Given the prior demonstrated benefit of semantic and syntactic information for word processing (e.g., Miller et al., 1951; Wright & Garrett, 1984), we predicted that having the words of the sentence stem randomized should harm accuracy and speed of decoding for the unscrambling task. There is some evidence from Schriefers, Friederici, and Rose (1998) that scrambled sentences can be used to predict the final words of a sentence. It was unlikely, however, that this would be the case in the present research because Schriefers and colleagues used only sentence stems that were three words long, and our sentences were generally longer. Other studies using more complex scrambling methods such as longer sentences (Simpson, Peterson, Casteel, & Burgess, 1989) and additional replacements (O’Seaghdha, 1989) have not found any priming benefits for scrambled sentences. Thus, we predicted that decoding of the scrambled word should be harder, and slower, in sentences for which the word order was jumbled.

In addition to a sentence providing context for an upcoming word, the sentence itself could predict that a specific word will be seen. In Experiment 2, the importance of the sentence was explored by manipulating the predictability of the final scrambled word. This level of predictability was previously normed by a separate set of individuals. Although our participants were only instructed to unscramble the final word and were never told to read the entire sentence, it was hypothesized that they should find it easier to unscramble the highly predictable than the unpredictable words (as in lexical decision studies such as Schwanenflugel & Shoben, 1985). We predicted that having a context in which the scrambled word met the built-up expectations of the reader should make the task of unscrambling the word easier because the reader may already have an idea of what to look for in that set of letters.

Overall, we predicted that target words using the fixed scramble type would be easier and faster to decode than those in the random scramble type (Experiments 1 and 2). Additionally, a more predictable target word should be easier and faster to decode. This predictability could stem from the orderliness of the sentence stem (Experiment 1) or the predictability of the target word itself (Experiment 2).

Experiment 1

Method

Participants. Thirty-four college students (19 women) aged 18–22 at a small regional university participated. A power analysis using G*Power (Faul, Erdfelder, Buchner, & Lang, 2013) on a repeated-measures Analysis of Variance (ANOVA) with two independent variables determined that the required sample size is 34. All participants were native English speakers who either volunteered to participate without compensation or received course credit. One additional participant was removed from analyses because the criteria of being able to correctly complete half or more of the trials was not met.

Materials. On each trial, participants were presented with one of 80 sentences in which the final target word was scrambled. The sentences (including the final scrambled word) were between four and eleven words in length (\(M = 6.2\) words). The 80 final target words were all nouns and had a frequency between 1,000 and 3,000 out of one
million according to the Corpus of Contemporary American English (COCA; Davies, 2008).

Across sentences, two variables were manipulated: the type of scramble used and the order of the preceding words in the sentence. For type of scramble, the final target word could be scrambled such that the first and last letters of the word were maintained (fixed scramble) or such that all letters were randomly scrambled (random scramble). For example, the target word emergency could be viewed as ergemceny (fixed scramble) or germecyne (random scramble). No other criteria were used when creating the scrambled words, but we attempted, as much as possible, to limit how often letters that appeared in order in the word also appeared together in the scrambled versions. For sentence type, the preceding words in the sentence were either in correct syntactic order (fixed order) or were randomly reordered (random order). In the random order sentences, we limited the reordering such that the random order never had more than two words in a row that were correctly placed in relation to each other. The manipulation of these two variables led to four possible sentence conditions. Italicization is used to highlight the target word in the following examples, but they were not presented in italics to participants. For example, the sentence, “The cable went out because of the horrible storm,” was seen as one of the following: “The cable went out because of the horrible sortm” (fixed scramble–fixed order), “The cable went out because of the horrible tsrmo” (random scramble–fixed order), “Went horrible because out cable the of the sortm” (fixed scramble–random order), or “Went horrible because out cable the of the tsrmo” (random scramble–random order).

Participants saw 20 sentences in each of these four conditions in a randomized order, and trial type was counterbalanced across participants. Each block of 20 trials had an equal number of sentences from each of the four conditions.

Procedure. Following approval by the DeSales University institutional review board, participants were recruited and tested individually. They were told that they would be viewing sentences presented on a computer screen one at a time using the program PsychoPy (Peirce, 2007). Their task on each trial was to unscramble the final word in the sentence as quickly and accurately as possible. After viewing each set of 20 sentences, participants could take as long a break as they desired. Upon decoding of the final target word in the sentence, participants used the keyboard to type the word in the correct spelling and hit the enter key to indicate completion. PsychoPy does not allow participants to fix typed errors, so participants were told to just keep going if they made a mistake or if they realized part way through that they were incorrectly unscrambling the target word. The timing of both the first keystroke of the typed word and of the enter key were recorded as measures of speed of unscrambling (timing began for each trial when the sentence first appeared). Although participants were encouraged to complete every trial, they were told that, if they were certain they would be unable to unscramble the word, then they could skip to the next sentence by just pressing the enter key. Only data from participants who accurately completed at least half of the unscramblings were included in the analyses.

Results
Scoring. Because participants were asked to type their answers, there were, unsurprisingly, some typographical errors. Any trial without a perfect match to the target word was inspected to determine whether this was an inability to complete the trial, an error in decoding, an error in typing, or a spelling error. For example, eight participants incorrectly spelled prescription as perscription, an unsurprising spelling error. Other obvious typographical errors included avariables (extraneous letter before the start of the word) and bariables (transposition by one position on the keyboard for the first letter) instead of variables. Any trial that was an obvious spelling or typing error was scored as correct unscrambling. These types of errors occurred on 274 of the 2,720 total trials and accounted for 13% of trials that were counted as correct. Clear mistakes such as pakenum for makeup were counted as incorrect as were trials where the participant was unable to make a guess. Additionally, two participants discovered that silence could alternatively be unscrambled as license. These two trials made up less than 0.08% of the trials and were marked as accurate.

Overall accuracy. To determine whether sentence or scramble type influenced accuracy, a two-way repeated-measures ANOVA was conducted with sentence type (fixed order or random order) and scramble type (fixed scramble or random scramble) as factors and accuracy as the dependent variable (see Figure 1). This revealed significant main effects of both scramble type, $F(1, 33) = 141.04, p < .001, \eta^2 = .81$, and sentence type, $F(1, 33) = 11.108, p < .005, \eta^2 = .25$, on accuracy. Overall
accuracy for fixed scramble (91.8%) was greater than for random scramble (62.0%), and accuracy for fixed sentence order (80.2%) was greater than for random sentence order (73.5%). There was a significant interaction effect between scramble type and sentence type, $F(1, 33) = 7.83$, $p < .01$, $\eta^2 = .19$. Post-hoc analyses using paired-samples $t$ tests with Bonferroni corrections demonstrated that accuracy was higher for fixed scramble than for random scramble for both sentence types ($p < .001$). Although accuracy was higher for fixed sentence order than for random sentence order when the target word had a random scramble ($p < .01$), there was no difference when the target word had a fixed scramble ($p = .99$).

**Speed of unscrambling.** Only trials where the word was correctly decoded were included in the analyses for speed of unscrambling. Using those trials, a two-way ANOVA examined the effect of scramble and sentence type on speed of unscrambling completion (see Figure 2). Speed of unscrambling was determined by measuring both the first letter typed ($FirstClick$) and when the participant hit the return key to move on to the next trial ($MoveOn$). Results for both measures did not differ qualitatively and thus only $MoveOn$ will be reported. For $MoveOn$, there was both a significant main effect of scramble type, $F(1,33) = 75.23$, $p < .001$, $\eta^2 = .70$, and sentence order, $F(1,33) = 51.89$, $p < .001$, $\eta^2 = .61$. Overall average completion speed for fixed scramble (7.8 seconds) was faster than for random scramble (14.0 seconds), and average completion speed for fixed sentence order (8.1 seconds) was faster than for random sentence order (11.1 seconds). There was also a significant interaction effect, $F(1,33) = 6.74$, $p = .014$, $\eta^2 = .17$. Post-hoc analyses using paired-samples $t$ tests with Bonferroni corrections demonstrated that participants were faster for fixed scramble than for random scramble for both sentence types ($p < .001$). Although participants were faster for fixed sentence order than for random sentence order when the target word had a random scramble ($p < .01$), there was no difference when the target word had a fixed scramble ($p = .28$).

**Discussion**

The results of Experiment 1 added to a body of literature suggesting that not all manipulations of words are equally easy to decode. When the first and last letters of the scrambled word were held constant, participants were faster and more accurate at decoding the word when the letters were completely scrambled. We found that the fixed scramble was easier to read regardless of the prior context of the sentence. This provides additional evidence for the claim that the middle of a word is less critical for identification than is the beginning (Rayner & Pollatsek, 1989) or the end of a word (Perea & Lupker, 2003). We also demonstrated that the context in which a scrambled word is presented was central to identification. Correctly ordering the preceding words of the sentence provided participants with extra contextual information.
that improved accuracy and speed of scrambled word decoding. However, participants only showed a benefit of the correctly ordered sentence when the word to decode had the random scramble. Of particular note is that participants were not required to read the entire sentence, they were merely instructed to decode the one scrambled word. Because the scrambled word was always the final word in the sentence, they did not ever need to pay attention to any other part of the text. Any use of the preceding sentence to aid with word identification was entirely driven by participants themselves.

The fact that the sentence order only influenced performance for the random scramble could suggest that, when the scramble condition was more difficult (as can be seen by the main effect of scramble type), participants were more likely to turn to the sentence itself for help in unscrambling the word. For the easier (i.e., fixed) scramble condition, the prior context was not important. This suggests that participants chose to take advantage of the context of the sentence when it was most beneficial for them to do so.

**Experiment 2**

In Experiment 1, we explored the value of sentence context for scrambled word identification by manipulating the order of the preceding words in the sentence. This meant that the final word either came at the end of a relevant sentence or after a seemingly random set of words. It could be argued that this is similar to having the target word in isolation as compared to having it in a sentence. Another way of exploring the role of context is to consider situations where the unscrambled word is, or is not, highly likely given the preceding context. In Experiment 2, we first used a norming procedure to identify a set of sentences that highly predicted a target word and another set of sentences that did not predict the target word. Then we presented those sentences to participants who completed a descrambling task. We predicted that highly predictable scrambled words (those that clearly match the previous context of the sentence) should be easier and faster to decode than unpredictable words.

**Method**

**Participants.** Fifty-seven adults were recruited using Amazon’s Mechanical Turk for the target word norming task. All “workers” self-reported as being both 18 or older and native English speakers. The length of time to complete the survey and hourly pay varied across participants depending on the version of the sentences they viewed and the speed at which they worked on the task. Data were collected over three postings, and the average time for completion of this task ranged from 11 min 54 s to 35 min 23 s (depending on the posting), and from this, participants were paid at an hourly rate between $3.60 and $7.56 based on their speed of task completion.

Thirty-six college students (29 women) aged 18–37 at a small regional university participated in the unscrambling task. All were native English speakers. Eight additional participants were removed from analyses because they did not follow instructions (1), did not meet the criteria of being able to correctly complete half or more of the trials (1), or because of computer error (6). Participants either volunteered to participate without compensation or received course credit.

**Materials.** Following exempt status determination from the DeSales University institutional review board, participants in the norming task were presented with sentences through Amazon’s Mechanical Turk. The questions were hosted on the online survey program Qualtrics (https://www.qualtrics.com/). Participants’ task was to rate the predictability of the final word of each sentence presented to them on 6-point Likert-type scale from 1 (very unexpected) to 6 (very expected). For each sentence, the final word was presented in all caps to ensure that participants were evaluating the predictability of the correct item. For example, we expected the sentence “Lavinia auctioned off the expensive JEWELRY” to be given a higher average rating than “Lavinia auctioned off the expensive SURFACE.” This goal of the norming procedure was to create the stimuli that would be used in Experiment 2.

An initial group of 23 workers viewed three versions each of 60 sentences for a total of 180 sentences. Each sentence stem was paired with a final word that was expected to have a high predictability rating; a final word that was expected to have a low predictability rating, and one that was expected to be neither highly predictable nor highly unpredictable (“average” ratings). Based on their average ratings for these sentences, a subset of the sentences was selected such that the predictable ending had an average rating of at least 4 out of 6, and the unpredictable ending had an average rating of less than 3 out of 6. Although we had initially hoped to have three levels of predictability, we were not able to create three distinct groups, and thus
the “average” ratings were not pursued further. Sentence stems that failed to find either a high or low predictability ending were given new final words in the second posting. This posting contained 61 sentences and was completed by 12 participants. The same procedure was followed to select high and low predictability endings. Because of the relatively small number of participants who completed the second posting, a final 22 participants viewed a set of 63 sentences, which included both the endings with a rating above 4 or below 3 from the second posting, and additional options for sentences that had not yet found an acceptable ending. These three postings resulted in a final set of 60 possible sentence stems with both a high and low predictability ending. As a result, the selection of an ending being either high or low predictability resulted from ratings from between 22 and 34 individuals.

After selection of the 60 sentence stems, it was discovered that a subset of the target words had multiple scrambles (such as tapas for pasta or below for elbow). For this reason, 48 of the sentence stems (without multiple unscramblings) were chosen to be analyzed as targets in the following experiment, and the 12 additional sentences were used as fillers. For these 48 target sentence stems, each had one high predictability and one low predictability ending. All sentences had between five and nine words, and the final word had between five and nine letters. Some examples for high/low predictability words, and the final word had between five and nine letters. Examples for high/low predictability words include, “Bryn drove too fast around the CURVE/GROUND,” “Her mother planned the extravagant WEDDING/ACCOUNT,” and Cassius hung the heavy PAINTING/NEWSPAPER.”

The average frequency of the final words as determined by the Corpus of Contemporary American English (COCA; Davies, 2008) for the high predictability and low predictability endings was compared. A two-tailed t test found that frequency level for the high predictability words (average 39,329) and low predictability words (average 32,107) did not differ, t(94) = 0.51, p = .61. Similarly, a two-tailed t test found that the average number of letters in the high predictability words (average 6.81) and low predictability words (average 7.0) did not differ, t(94) = -.72, p = .47. Importantly, however, the high predictability endings did have a significantly higher rating (average 4.93) than the low predictability endings (average 1.89), t(94) = 34.02, p < .0001, d = 6.95 one-tailed.

Procedure. The procedure for Experiment 2 was nearly identical to that in Experiment 1. On each trial, participants were presented with one version of each of the 48 previously normed sentences and 12 filler sentences.

For each of the 48 target sentences, either the high predictability or the low predictability ending was used, and as in Experiment 1, the final word could have a fixed or random scramble. For example, the sentence stem “The little girl thanked the kind” could have the final word be woman (high predictability) or turkey (low predictability). This sentence was seen as one of the following: “The little girl thanked the kind woman” (fixed scramble–high predict), “The little girl thanked the kind anomow” (random scramble–high predict), “The little girl thanked the kind tkurey” (fixed scramble–low predict), or “The little girl thanked the kind vekeyu” (random scramble–low predict). As in Experiment 1, participants were given a break after each set of 15 sentences, and their instructions were the same as before.

Results

Scoring. As in Experiment 1, answers that were clearly typos or spelling errors were scored as correct. Spelling or typing errors occurred on 121 of the 1,728 total trials and accounted for 10.7% of trials that were counted as correct.

Overall accuracy. To determine whether predictability or scramble type influenced accuracy, a two-way repeated-measures ANOVA was conducted with scramble type and predictability as factors, and accuracy as the dependent variable (see Figure 3). This revealed significant main effects of both scramble type, F(1, 35) = 159.42, p < .001, η² = .82, and predictability, F(1, 35) = 298.83, p < .001, η² = .90, on accuracy. As in Experiment 1, overall average accuracy for fixed scramble (80.9%) was greater than for random scramble (50%). Additionally, average accuracy for the high predictability words (80.0%) was greater than for low predictability words (50.9%). There was also a significant interaction effect, F(1, 35) = 10.97, p = .002, η² = .24. Post-hoc analyses using paired-samples t tests with Bonferroni corrections demonstrated that accuracy was higher for fixed scramble than for random scramble for both levels of predictability (p < .001). Accuracy for high predictability words was greater than for

---

3Although data from the 12 filler trials was not included in the following analyses because of the complication that some words had multiple unscramblings, we did examine these trials. The filler trials showed the same pattern of results as the 48 target sentences. It is unlikely, therefore, that participants were aware of the differences between the target and filler sentences.
low predictability words for both scramble types ($p < .001$). Overall, accuracy was highest for fixed scramble with high predictability and lowest for random scramble low predictability.

**Speed of unscrambling.** Only trials where the word was correctly decoded were included in the analyses for speed of unscrambling. Using those trials, a two-way ANOVA examined the effect of scramble type and predictability on speed of unscrambling completion (see Figure 4). Results for the two measures of speed did not differ qualitatively, and thus only *MoveOn* (when participants indicated that they had completed typing the word) will be reported. For *MoveOn*, there was both a significant main effect of scramble type, $F(1,35) = 44.40, p < .001, \eta^2 = .56$, and predictability, $F(1,35) = 34.23, p < .001, \eta^2 = .49$. Once again, overall average completion speed for fixed scramble (8.3 seconds) was faster than for random scramble (19.7 seconds). Additionally, average completion speed for the high predictability words (9.4 seconds) was faster than for low predictability words (18.7 seconds). There was also a significant interaction effect, $F(1,35) = 12.64, p < .005, \eta^2 = .27$. Post-hoc analyses using paired-samples $t$ tests with Bonferroni corrections demonstrated that participants were faster for fixed scramble than for random scramble for high predictability words ($p < .05$) and low predictability words ($p < .001$). Although participants were faster for high predictability words than for low predictability words when there was a random scramble ($p < .001$), there was no difference when the target word had a fixed scramble ($p = .36$).

**Discussion**

Once again, we found that participants were fastest and most accurate at unscrambling target words when the first and last letters were held constant. This was true regardless of the level of predictability of the final words. This served as additional support for our initial prediction that the beginning and end of scrambled words would be particularly important for scrambled word identification. Additionally, we found that participants were more likely to be able to decode the high predictability words than the low predictability words, regardless of the type of scramble. Scramble type and predictability interacted such that high predictability words with a fixed scramble were the easiest to read, and low predictability words with a random scramble were the most difficult to read. This greater facility for predictable words suggests that the previous words in each sentence built up an expectation about what that final word might be. As demonstrated by Van Berkum et al. (2005), it is possible that, by the end of a sentence, readers had focused in on a small set of possible final words that they were considering. When the scrambled words aligned well with the context and matched one of those possible words, this expectation might have made the words easier to identify because readers only needed to sample from that small lexical subset in order to complete the task. The predictability of a word has been found to influence response times...
in both lexical decision (Wright & Garret, 1984) and word naming tasks (West & Stanovich, 1986). However, our participants only showed a reduced response time for highly predictable words when the random scramble type was used. When the easier (i.e., fixed) scramble condition was used, the predictability of words did not significantly influence speed of response. Again, this shows that participants were able to take advantage of the contextual information in each sentence rather than just focusing on the target word, although they might have only chosen to do so when faced with a more challenging scramble condition.

**General Discussion**

Prior work has demonstrated that there is a time cost to reading words with reordered letters (Rayner et al., 2006), but most studies of scrambled word recognition have focused on the value of that word for priming tasks (e.g., Perea & Lupker, 2003) rather than reading in context. Our goal was to more closely examine factors that influence people’s ability to decode scrambled words. We found that both the method of word scrambling and the prior context of the sentence significantly impacted accuracy and speed of scrambled word decoding, but not to the same degree.

Across Experiments 1 and 2, we manipulated the type of scramble used. It has previously been demonstrated that the beginning and the end of a word are particularly important for word identification. When letters are transposed or substituted, alterations that occur at the beginning of a word lead to slower overall reading times (Rayner & Kaiser, 1975; Rayner et al., 2006), and transpositions at the end of a word inhibit priming effects (Perea & Lupker, 2003). In line with the literature, participants were more likely to be able to accurately unscramble the final word of the sentence—and did so more quickly—when the first and last letters were in the correct position than when they were not. This was true regardless of the sentence level manipulations used. As previously demonstrated, these results indicate that the first and last letters of a word are important not only for letter transpositions but also for complete scramblings. We also provide a contrast to Rawlinson’s (1976) finding that scrambled text does not negatively impact comprehension. Although we did not directly measure comprehension, we did find that the way in which a word is scrambled influences not only speed of decoding but also whether a word is even identifiable at all. We can reasonably assume that when a word could not be recognized, the broader comprehension of the sentence did, in fact, suffer. These differences in results may have stemmed from the stricter scrambling method used in the present studies than in Rawlinson’s study. Perhaps the fact that Rawlinson held the first two and last two letters of the word constant was enough information to allow the reader to easily interpret the word, thus not impairing comprehension. As speed for reading the target word was not measured in that study, however, it is unclear whether that scrambling approach harmed reading time even if it did not harm comprehension. These results generally demonstrate that not all word scramble manipulations are equally problematic.

Our finding that most scrambled words can be identified would support any word recognition model that allows for some letter position flexibility (such as the spatial coding or SERIOL model). However, the fact that the fixed scrambling method was less disruptive overall to reading provides further evidence in support of the predictions of word recognition models, such as the spatial coding model (Davis, 2010), that give extra weight to the external letters of a word for the purpose of identification. This distinction should be considered in future modifications to word recognition models. Our results suggest that these positions in the word provide an important cue for word identification, possibly by narrowing the scope of possible words. By providing the first and last letters, we significantly decreased the set of lexical items from which the scrambled word could be found. This, of course, only helped if participants took advantage of this extra information.

The second factor that we examined was the role of the context in which the scrambled words were presented. In Experiment 1, the scrambled words were always found at the end of the sentence, but the usefulness of previous words was sometimes limited by having them in a random ordering. In Experiment 2, the predictability of the final word was manipulated. Sentences that are correctly ordered have been found to be easier to process than those with a seemingly random set of words (Miller & Isard, 1963) or even those that are acceptable but do not follow canonical word order (Tanaka et al., 2007). The assumption is that prior words in a sentence provide a context that then leads to easier recognition of individual words. Given that recognition of the final word of a sentence is faster and more accurate when that lexical item is expected (West & Stanovich,
1986; Wright & Garret, 1984), we predicted that both the correctly ordered sentence condition (Experiment 1) and highly predictable final word condition (Experiment 2) would lead to fast and accurate unscrambling. Although this prediction was generally confirmed, sentence context did interact with scramble type. In Experiment 1, we found that having a correct sentence order only improved accuracy and speed of unscrambling for the random scramble trials. In Experiment 2, having the target words be high predictability always improved accuracy, but it only improved reaction time for the random scramble trials. Across both experiments, when the fixed scramble type (which had an overall higher accuracy rate) was used, contextual information did not influence response times. It is possible that, when given a more difficult unscrambling (in this case the random scramble), readers may need to make use of any available predictive information in the sentence stem to help them complete the task. If the sentence then provides no useful context, readers need to rely on just their unscrambling ability (such as it might be for a single word with no context) or take extra time to unscramble the sentence stem. Help from the sentence context might not be as necessary for the easier scramble type. Although these results overall support previous literature showing that predictive context may be used to help identify an upcoming word, we show here that context is not equally effective across all sentences, but that it is most beneficial in particularly difficult decoding situations. In the present research, when given altered text, readers appeared to focus first on the target word and then only looked further, considering context, when necessary.

Although the discussion of word recognition models thus far has only focused on individual words, any theory that attempts to explain word recognition in the larger context of a sentence will need to consider the fact that information at both the individual word and sentence level matter for identification but perhaps not, as we demonstrated here, to the same degree. The complexity of the scramble method for a word may determine the degree to which context is used. It is possible that this same effect may occur for other types of altered, or otherwise difficult to read, words. Word recognition models that go beyond the individual word to consider the phrases or sentence level should consider the relative importance of these cues. In addition to informing models of word recognition, a deeper understanding of the relative importance of specific letter positions and context cues on people’s ability to interpret words could be useful in a more practical setting. For example, this knowledge might be beneficial for better understanding the broader reading process, the reading difficulties of young readers, or even in explaining effects of developmental or acquired dyslexia. For example, one rare form of acquired dyslexia causes individuals to have difficulties with letter position encoding. As a result, they may flip the location of letters within a word, thus reading *forth* as *froth*. Interestingly, these migrations are much less common for the first and last letters of a word than for the internal letters (Friedmann & Gvion, 2001). Our findings add to an understanding of how the reader responds to internal as compared to external alterations in letter position. This knowledge may add in the creation of word recognition models that can more accurately predict this form of letter position dyslexia.

**Limitations and Future Directions**

Possible limitations with the design of the present studies should be taken into account when considering the implications of this work. One of the downsides of the program we used to present the stimuli is that participants were not able to see what they were typing, nor were they able to fix typing errors. As a result, there were some situations where the accuracy of the unscrambling was not entirely obvious. For example, we had to determine whether *vessle* was either (a) a misspelling of *vessel*, (b) an unintended translation of the last two letters during typing, or (c) the result of a participant not being able to unscramble the word and randomly typing letters (and getting very close by chance). Across both experiments, a total of 395 trials had errors that were determined to be typing or spelling errors. These were distributed across participants with only one participant making no such mistakes. Of these judgments, the vast majority (89.1%) were cases where the participant clearly knew the correct word but made errors. For example, the participant might have started with an error but ended correctly, incorrectly pluralized, had an additional extraneous letter, or used a common misspelling. There were only 43 trials total (1.6% of all trials scored as accurate across both experiments) that were less clear and could be up to interpretation. Because of the rarity of these cases, any mistakes on our part when classifying the typos were unlikely to have any significant effect on our analysis. However, it would be preferable to use a data collection
method where participants could fix errors, and thus fewer guesses would have to be made. For this reason, future studies may wish to use more flexible experiment building software packages such as PsyToolkit (Stoet, 2017).

Given the finding that overall reading speed may moderate how word alterations influence priming tasks (Duñabeitia et al., 2014), it would be interesting to know whether this is also true for unscrambling tasks. One limitation of the present study is that, although we only included native English speakers in our tasks, we did not have any independent measures of reading ability. This could be measured through simple reading time (Duñabeitia et al., 2014) or by examining reading comprehension. Future studies may wish to examine whether the benefits of a fixed scramble condition are equally large for more or less proficient readers.

One problem that we encountered was that some of our target words had more than one possible unscrambling. For this reason, we were not able to analyze all the possible trials in Experiment 2. This limitation, however, could lead to an interesting line of enquiry. We demonstrated in the current work that the way a word is scrambled and the context in which it is found both impact word identification. Although we found that scramble type had a larger effect than sentence context overall on performance, it is not entirely clear how these factors play their role. One possibility would be to identify words with more than one possible unscrambling. If the target word was in the fixed scramble position for one unscrambling, but the sentence structure predicted the other unscrambling, what would our participants respond? This manipulation would allow an additional way for us to directly compare the importance of these cues.

In Experiment 2, we examined the effect of predictability on the final word unscrambling. Although the set of words used in both the high and low predictability conditions did not differ overall on word length or frequency, each list had a different set of words. In an ideal situation, the same scrambled word would be used with two separate sentence frames so that each target word could be, across participants, seen as either high or low predictability endings. In the current work, we compared sentence endings that were given either very high or very low predictability ratings and found, as expected, that the more predictable endings were easier to unscramble. It would be informative, however, to know whether this is an all or nothing effect (i.e., predictable or not) or whether it is a graded effect. Perhaps there is a threshold for how likely the final word is before we see any benefit for accuracy of identification. Future studies that compare words across a wider range of predictability (perhaps at three levels) would help to explore this question.

Although we demonstrated that the first and last letter of a word are differentially important for word identification, it is possible, given our manipulation, that there is a second explanation for these results. Because we compared target words with the first and last letters maintained versus those that were completely scrambled, these two conditions were not exact comparisons as far as the number of letters to be rearranged. For example, in a six-letter word, you would only need to reorder the middle four letters instead of potentially all six (although the first and last letters in the random condition could not be in the accurate position, it was possible for one or more of the other letters to, by chance, be in the correct location). This could mean that participants performed better on the fixed trials because they are easier due to the number of letters to unscramble. The fact that only half of the trials had target words in this condition means that it is unlikely that participants were expecting and, thus, taking advantage of this possible benefit, but it should be considered. To confirm that it is the position of the first and last letters in particular that are important, additional studies should compare performance on those words to words where the middle two letters are fixed. This would allow for a more matched comparison set. As a related question, are both the first and last letter positions necessary for this effect? Would just holding the first letter constant also improve word identification?

There are many other open questions regarding word identification. We always put our scrambled word at the end of the sentence so as to potentially build up context and expectations. Our instructions to participants, however, did not tell them to read the whole sentence. If the scrambled word came first, would they still be likely to use the other information available? Another issue concerns the words that we asked participants to read. One commonality across much of the literature exploring word recognition is a focus on nouns. How does people’s ability to decode scrambled nouns compare to performance for other parts of speech? Would people see the same effects of context and scramble type in that case?

In summary, we examined two factors that

In summary, we examined two factors that
influence people’s ability to decode scrambled words. We found that holding the first and last letters constant made the scrambled words easier to identify. This adds to a body of literature suggesting that the first and last letter positions of words are particularly important for word identification. Although words where the first and last letters were not held constant were more difficult to unscramble overall, we found that context could influence performance on those trials. Decoding accuracy and speed improved when the sentence helped to predict the identity of the scrambled word. Having the prior words in the sentence in the correct order (versus randomly scrambled), or having the target word itself be a likely ending to the sentence, increased accuracy and speed of decoding.

References


Author Note. Sarah J. Starling, Department of Social Sciences, DeSales University; Kelsey Snyder, Department of Social Sciences, DeSales University.

The authors wish to thank John Blaise and Serena Nyan for help with data collection; William Guido for helpful conversations about the analyses; and Holly Nees, Daniel Rybak, and Lindsay Romanic for their work on a pilot study that helped inform the design of these experiments. Special thanks to Psi Chi journal reviewers for their support.

Correspondence concerning this article should be addressed to Sarah J. Starling, Department of Social Sciences, DeSales University, Center Valley, PA 18034. E-mail: Sarah.Starling@desales.edu
The first interaction with a potential romantic partner is a crucial moment. The impressions that form from these first interactions can have a lasting impact on the social dynamic. Researchers have demonstrated that the formation of first impressions is an automatic process, sometimes occurring without conscious awareness (Todorov & Porter, 2014). These initial evaluations then influence later judgments about an individual (Kahneman, 2011). Expectations regarding a social interaction begin to form at first sight. The physical characteristics of an initiator, including facial cues, are an important component of the preliminary evaluation of an initiator and can be a strong determinant in a recipient’s willingness to pursue further interaction (Schröder-Abé, Rentzsch, Asendorpf, & Penke, 2016). Another important determinant of the expectations for a social interaction is the opening line used to initiate the interaction. Bale, Morrison, and Caryl (2006) suggested that opening lines are important because they can serve as a form of display, providing information about the qualities that an initiator possesses. Because social interactions are influenced
by both the initial visual and verbal interactions, the present study explored the interaction between facial attractiveness and expected opening line usage.

Conversation initiators often rely on opening gambits, informally known as pickup lines, to begin social interactions. Opening gambits are typically viewed as being direct, innocuous, and flippant (Cunningham, 1989). Direct opening gambits are statements or questions that explicitly demonstrate an individual’s intention to a recipient such as, “Hey beautiful, how are you today? Would you like to go out for a cup of coffee?” Innocuous opening gambits are statements or questions that could be considered open ended. For example, “Hey, I’m new around here, where’s your favorite place?” Innocuous gambits typically do not directly display affection or romantic intent, but they do display one’s interest in engaging a recipient in conversation (Lewandowski, Jr., Ciarocco, Pettenato, & Stephan, 2012). Finally, flippant opening gambits are the stereotypical pick-up lines that could be interpreted as cute or funny (Lewandowski et al., 2012). An example of a flippant pick up line is, “Do you know how much a polar bear weighs? Enough to break the ice. Hi, I’m (insert name here).” (For additional examples of opening gambits, see the Appendix).

Although the choice of opening gambit provides information about an initiator, mate selection information is provided even before the first words are spoken as a recipient evaluates the attractiveness of the initiator. Attractiveness is often considered a socially constructed quality assigned to individuals based on their apparent traits and other’s perceptions of them. Within the context of the present study, being considered attractive would imply that an individual is visually appealing to others. Generally, attractiveness ratings are developed based on distinct physical features. Classic research has found that physically attractive individuals receive more favorable social encounters than individuals who are considered less attractive (Adams, 1977). Recent research has identified important social consequences of attractiveness such as higher career success rates, more romantic interactions, and others’ subjective views of the attractive individual as possessing other positive, high quality traits (Little, Jones, & DeBruine, 2011). More attractive people have higher ratings of self-confidence and often display an extroverted personality (Talamas, Mavor, & Perrett, 2016). Surprisingly, this extroverted trait can be determined after only 50 ms of exposure to a face (Talamas et al., 2016). Although many factors influence attractiveness, the current study focused specifically on facial attractiveness.

Facial attractiveness is dependent on certain facial cues and traits that are readily identified and analyzed by an evaluator, and the characteristics associated with an attractive face are generally agreed upon across cultures and individuals (Little, 2014). The discrimination of physical factors involved in the mate-selection process may trace back to increased reproductive fitness associated with success in finding mates fit to reproduce. Feminine facial qualities such as large eyes, smooth skin, and big lips are seen as attractive because these are indicators of female fertility (Meltzer, McNulty, Jackson, & Karney, 2014). Other facial features that have an evolutionary impact on mate selection are youthful attributes, weight cues, facial coloration, averageness, symmetry, and masculinity or femininity (Little, 2014). Within this evolutionary view, people value attractive qualities in mates primarily because these qualities act as indicators of good genetics. Individuals who possess these genes have an increased potential for reproductive success (Little, 2014). An initial evaluation of facial characteristics provides information regarding this potential reproductive fitness.

In addition to potential evolutionary benefits, attractive individuals have also been found to be more successful than less attractive individuals in inducing positive social encounters (Adams, 1977). This suggests that attractive individuals are perceived to possess personality traits or social behaviors that influence others around them, resulting in higher ratings of likability, agreeableness, and desirability. This tendency for a single observation to influence multiple judgments is an example of exaggerated emotional coherence, commonly referred to as the halo effect (Kahneman, 2011). The halo effect is a cognitive bias in which the overall opinion of one dimension of a person influences beliefs about other traits relative to that individual. This bias has the ability to influence a recipient’s perception of an approaching individual in a social encounter by exaggerating the reliability of evaluations. If a target perceives an approaching individual as attractive, based on facial characteristics, the opening gambit the recipient receives could be viewed as more desirable or attractive due to the initiator’s physical appearance. The role of expectations created by facial attractiveness is an important aspect in the
formulation of novel impressions, as well as the social norms of attractiveness and the success of intuitive analysis of an individual.

The facial attractiveness of a relationship initiator may be an influential factor in the development of expectations for the social interaction. Previous research has shown that the more attractive that individuals are perceived to be, the more often they are paired with a direct opening gambit (Cooper, O’Donnell, Caryl, Morrison, & Bale, 2007). Men who are perceived as more dominant and attractive than average have been found to prefer using direct style opening gambits (Ahmetoglu & Swami, 2012). Men who are perceived as less attractive are generally paired with either innocuous or flippant opening gambits (Ahmetoglu & Swami, 2012). Including facial attractiveness, the physical appearance of a woman can also influence a man’s choice of opening gambit style during a social encounter. Men are more likely to select a direct opening gambit for women who they perceive to be more physically attractive. In contrast, women with lower attractiveness ratings frequently receive innocuous or flippant style opening gambits (Wade, Butrie, & Hoffman, 2009). Along with perceived attractiveness, women who are seen as fit (e.g., having a slim body size) are seen as more attractive than those who are perceived as less fit (Swami et al., 2010). Women with lower attractiveness ratings frequently receive innocuous or flippant style opening gambits (Wade et al., 2009).

Men tend to be influenced by halo effects more often than women when rating attractiveness (Lorenzo, Biesanz, & Human, 2010). Within the halo effect, dialogue from attractive individuals could be seen as more competent and influential, simply due to their desirable physical characteristics. Lammers, Davis, Davidson, and Hogue (2016) found that, with a single headshot photograph, a person can decide whether someone is attractive or not. Adams (2012) demonstrated the impact of first impressions on a single head shot by cropping a picture to include just face and hair. This research found that desirable traits positively correlated with attractiveness ratings.

Direct opening gambits are not only associated with higher levels of physical attractiveness, but they are also found to be more effective in capturing a potential mate’s attention. Research has shown that men prefer more attractive women to use direct opening gambits when approaching them (Wade et al., 2009). Examples of lines men found more effective were “Want to go watch a movie tonight?” and “Want to go to a bar later?” Thus, results have demonstrated that direct opening gambits were more effective in capturing the attention of a potential mate, compared to the other opening gambit styles.

Although previous research has examined receptivity to lines used by people of differing levels of attractiveness, the current study set out to investigate the recipients’ expectations regarding the type of opening gambit that would be used by initiators with varying ratings of attractiveness. Expectations tend to shape the initial reactions to social interactions. Understanding how attractiveness influences a person’s expectations regarding how a person would initiate a social interaction is an important step in developing an understanding of receptivity to different styles of relationship initiation. The primary hypothesis for the current study predicted that more attractive individuals would be expected to use a direct style of opening gambit. A secondary hypothesis predicted that less attractive individuals would be expected to present an innocuous or flippant style of opening gambit significantly more than a direct style of opening gambit.

### Method

**Participants**

Participants involved in this study were 121 undergraduate students from the sponsoring university. Participants were recruited through social media sites such as Facebook, GroupMe, and Twitter, as well as an online participant pool hosted by Sona Systems. Sixty-three participants who were enrolled in psychology courses at the University of Central Arkansas registered for the study through the Sona participant pool. The remainder of the participants (n = 58) chose to participate through email or social media invitations.

Participants ranged from 18 to 51 years of age, with a mean of 20.90 years (SD = 3.62). Eighty-six participants were women (71.1% of the sample; 1 participant did not report gender). Thirty-two participants reported an exclusive or predominant sexual attraction to women, 86 reported an exclusive or predominant attraction for men, and two reported an equal attraction to men and women (1 participant did not report sexual attraction). The two participants reporting a mixed attraction were given participation credit without completing the opening gambit or stimulus ranking tasks due to uncertainty about which stimuli to present during the task. The only demographics collected were age, gender, and the participant’s self-reported sexual
Attractiveness and Opening Gambits | Wood, Charlton, Goodman, and Thompson

The 15 opening gambits used in this study were developed through a multistep process. This project was completed as part of an undergraduate research methods laboratory class. The first assignment for the 12 members of the class was to identify a list of opening gambits from personal experience. This activity resulted in an initial list of 95 opening gambits that the class members had either used themselves, heard others use, or had personally been presented with in real-world encounters. Rather than using existing opening gambits from the published literature, this approach was selected in order to generate a novel set of gambits that were modern and relevant to the participant pool, as well as to give the class the experience of generating items for a research study.

The 95-item list was trimmed down to the final 15 items shown in the Appendix through several steps. First, the research team separated into four groups of three people each to categorize each gambit as direct, innocuous, or flippant. All four groups agreed on the categorization of 72 of the 95 (76%) opening gambits on this first independent ranking. The groups then separately reviewed the 23 gambits that did not have a consensus type and recategorized these. Groups were not told how the other groups originally categorized these gambits. This second round of categorization resulted in agreement on 86 of the 95 gambits (91%). A final category determination was made by the entire research group, resulting in 100% agreement for the 95 opening gambits. Next, the four groups of students selected their top 10 gambits in each of the three categories. At this point, any gambit that was not identified in the top 10 for its category by at least one group was removed from further consideration. Finally, each of the 12 members of the class independently identified their top five gambits in each category using an online Excel sheet. The five gambits for each category that were most often selected in the top five grouping by the research class members were retained for use in this study. All 15 of these items had unanimous agreement during the first round of categorization. The list of opening gambits, and their classification type, are shown in the Appendix.

Facial stimuli. The photographs presented to participants during the study were chosen from the Chicago Face Database (Ma, Correll, & Wittenbrink, 2015). The Chicago Face Database provides standardized photographs of male and female faces of varying race and ages. Stimuli used for the study were selected based on standardized attractiveness scores and race. All stimuli selected showed a neutral facial expression because previous research supported the neutral or happy faces as being viewed as more attractive (Mueser, Grau, Sussman, & Rosen, 1984). Neutral and happy facial expressions were not shown to differ in physical attractiveness ratings in the Mueser and colleagues study (1984). Although the Chicago Face Database scores stimuli on a variety of dimensions, only the attractiveness ratings of male and female models between the ages of 19 and 26 were evaluated for this study. The research team selected five photographs of women rated more attractive and five photographs of women who were rated less attractive. The same method was replicated for the pictures of men used in the study.
Face Database were based on a 9-point Likert-type scale from 1 (unattractive) to 9 (highly attractive). For both the attractive rated pictures of women and men, the highest ranked stimulus for each race was selected. These ratings were between 3.8 to 5 for men and 4.7 to 5.4 for women. The less attractive stimuli were selected from the lowest Likert scaled scores on attractiveness for each race for both men and women. These ratings were between 1.7 and 2.2. The races of the models presented in the stimuli included three White women and men, three Black women and men, two Latino women and men, and two Asian women and men. The distribution of race for the racial stimuli was selected based on the demographics of the host university where the majority races are White and Black. Asian and Latino faces were included to ensure diversity of the sample stimuli. The age of the men selected was 23 to 26; the age of the women selected was 19 to 26. A 2 (gender of model) x 2 (high/low attractiveness) x 4 (race of model) Analysis of Variance (ANOVA) on the attractiveness ratings from the Chicago Face database found that the attractiveness of the stimuli differed in terms of gender (women: M = 3.61, SD = 1.59; men: M = 3.23, SD = 1.30), F(1, 19) = 10.53, p = .03, ηp² = .32, and attractiveness level (high: M = 4.80, SD = 0.50; low: M = 2.03, SD = 0.24), F(1, 19) = 567.49, p < .001, ηp² = .99. No difference was found as a function of the race of the stimuli, F(3,19) = 1.21, p = .41, ηp² = .48. The interaction between gender and attractiveness level was not significant, F(1,19) = 1.69, p = .21, nor was the interaction between race and attractiveness level, F(1,19) = 3.32, p = .14, ηp² = .21. These statistical differences confirmed that the low and high attractiveness stimuli did differ in terms of attractiveness and that there were no systematic differences across the four races represented.

A similar 2 (gender of model) x 2 (high/low attractiveness) x 4 (race of model) ANOVA explored differences in age across the stimuli. This test found no difference in age across gender F(1,19) = 0.44, p = .54, ηp² = .10, attractiveness level F(1,19) = 1.91, p = .24, ηp² = .32, or race F(3,19) = .89, p = .52, ηp² = .40. The interactions between these variables were also not significant: gender by race, F(3,19) = 0.95, p = .50, ηp² = .42; gender by attractiveness level, F(1,19) = 5.24, p = .08, ηp² = .57; race by attractiveness level, F(3, 19) = 3.41, p = .13, ηp² = .72; and gender by attractiveness level x race, F(3,19) = 4.47, p = .09, ηp² = .77.

Opening gambit multiple-choice. Participants were shown either the five men or the five women, depending on the participants’ reported sexual preference. Because the stimulus type shown to participants was determined based on self-reported sexual preference and not the participants reported gender, we classified participant responses based on the target stimulus gender throughout the results and discussion. Each stimulus headshot was presented five times, each time with three different opening gambits, one of each opening gambit type (direct, flippant, and innocuous). The opening gambits and the opening gambit triplicates that were shown to participants are shown in the Appendix. Each of the 10 stimuli (5 attractive and 5 less attractive) was shown five times, for a total of 50 trials. Order of presentation was randomized across participants. Participants were asked to select the gambit they thought the person in the photo would most likely use.

Attractiveness rating. To verify that participants viewed the facial attractiveness of the facial stimuli differently across conditions, each of the 10 photographs within a sex was presented to participants with a Likert-type scale ranging from 0 (unattractive) to 5 (extremely attractive). Participants indicated the level of attractiveness that they thought best described the pictured individual. The full materials for this study are available for review/download from the Open Science Framework: https://osf.io/yl6zt/

Procedure
Participants completed the study in a location of their choosing. All materials were presented online. The survey required approximately 15 minutes to complete. Participants were supplied with an informed consent cover letter at the beginning of the study. Participants who accepted the informed consent were then asked to answer questions about their age, gender (woman, man, or other), and self-reported sexual preference (“I am exclusively or predominantly interested in women,” “I am equally interested in women and men,” and “I am exclusively or predominantly interested in men”). Participants then completed the opening gambit selection task followed by the stimulus ratings task. The procedures were the same for each participant regardless of gender or their reported sexual preference, with the only difference being that those who reported a primary interest in women were shown the pictures of women while those who reported a primary interest in men were shown the pictures...
of men. After the opening gambit selection task, participants completed the attractiveness rating task. The order of presentation of stimuli was randomized within both the opening gambit selection and the stimulus rating tasks. Once participants completed the survey, they were presented with a debriefing statement.

Results

A total of 121 women and men participated in our research study. We excluded five participants due to incomplete participation or reporting an equal sexual attraction to women and men. Eighty-three of the 86 women in the sample reported an exclusive or predominant sexual attraction to men (96.51%), one reported an exclusive or predominant attraction to women, and two reported an equal attraction to both. Of the 34 men who participated in the study, 31 reported an exclusive or predominant sexual attraction to women and three reported an exclusive or predominant attraction to men (91.18%). Two participants reported an equal preference between men and women. Participants’ reported sexual preference was not included as a variable in the analyses due to the small number of same-sex preference reported by participants (only 6 of the 121 participants, 4.96% of the sample).

Stimuli Ratings

To evaluate the effectiveness of the attractiveness intervention across facial stimulus type, a general measure of attractiveness was calculated by summing the ratings for the five pictures in the more attractive and less attractive categories. Possible total attractiveness scores for each participant ranged from 0 to 25. The mean total attractiveness score for the attractive pictures of women was 15.13 (SD = 5.37). The mean total attractiveness for the less attractive pictures of women was 3.97 (SD = 3.06). For the pictures of men, the mean total ratings for the attractive and less attractive stimuli were 14.64 (SD = 4.56) and 3.26 (SD = 3.24), respectively. Differences between these groups were evaluated using a 2 (attractive or less attractive; within-subject) x 2 (target stimulus gender, men or women; between-subject) mixed-measures ANOVA. This test found a main effect of picture type (attractive or less attractive), $F(1, 116) = 694.68, p < .001, \eta^2_p = .86$, but not a main effect of target stimulus gender, $F(1, 116) = 0.68, p = .41, \eta^2_p = .006$, or interaction between these variables, $F(1, 116) = .07, p = .79, \eta^2_p = .001$. Results displayed in Figure 1 show the clear separation in attractiveness ratings across the attractive and less attractive stimuli types.

Opening Gambits and Attractiveness

The main hypotheses for this study were evaluated using three repeated-measures ANOVAs, one for each opening gambit type (flippant, direct, and innocent). The separate analyses were required due to the dependent nature of the gambit selection activity. Because participants were forced to select one of the three gambit types on each of the 50 questions, omnibus tests including stimulus type (attractive or less attractive), stimulus attraction (men or women), and participant gender (woman or man) always resulted in a nonsignificant difference due to the identical means across the levels of these groups. All follow-up testing used Bonferroni corrected pairwise comparisons.
Hypothesis 1
The first hypothesis stated that attractive individuals would use direct opening gambits more than less attractive individuals. This hypothesis was evaluated using a repeated-measures ANOVA with target stimulus gender (men, women; between-subject) and stimulus type (attractive, less attractive; within-subject) as the independent variables and number of selections of the direct opening gambit as the dependent variable. The mean number of times the direct line was selected for the attractive stimuli ($M = 10.84, SD = 3.33$) compared to the number of times direct lines were selected for the less attractive stimuli ($M = 5.94, SD = 2.87$) was significantly different, $F(1, 114) = 72.97, p < .001, \eta^2_p = .39$. These differences are shown as the far right bars on Figure 2.

The difference in expected usage of direct opening gambits by the pictured men ($M = 8.86, SD = 1.70$) differed significantly from the mean expected usage of direct lines by the pictured women ($M = 7.10, SD = 2.16$), $F(1, 114) = 21.02, p < .001, \eta^2_p = .16$. The interaction between stimulus type and target stimulus gender was also statistically significant, $F(1, 114) = 14.34, p < .001, \eta^2_p = .11$. The right pair of bars on the top (attractive) and bottom (less attractive) panels of Figure 3 show this interaction.

A series of $t$ tests explored the significant interaction between target stimulus gender and attractiveness. Two dependent-samples $t$ tests compared expected usage for attractive compared to unattractive men and for attractive compared to unattractive women. Expected usage of direct gambits by attractive ($M = 8.23, SD = 4.04$) and unattractive ($M = 5.97, SD = 3.38$) women differed significantly, $t(84) = 2.07, p = .047, d = 0.53$. A significant difference was also observed between attractive ($M = 11.79, SD = 2.44$) and unattractive ($M = 5.93, SD = 2.68$) men, $t(84) = 14.09, p < .001, d = 2.29$. A visual comparison of the mean difference in expected gambit usage by attractiveness suggests that level of attractiveness had a larger effect on the expected usage of direct gambits by the pictured men than it did the pictured women.

Two independent-samples $t$ tests evaluated the difference in expected direct gambit usage between the pictured men and women for each of the attractiveness types. The expected usage of direct opening gambits by the pictured men and women did not differ for the less attractive stimuli (women: $M = 5.97, SD = 3.38$; men: $M = 5.93, SD = 2.68$), $t(114) = .06, p = .95, d = 0.01$. For the attractive stimuli, the mean expected usage of direct opening gambits for both men and women were higher than that observed for the less attractive stimuli (as indicated by the dependent measures $t$ tests described in the preceding paragraph), but this difference was much greater for the men than the women (men: $M = 11.26, SD = 2.44$; women: $M = 8.23, SD = 4.04$), $t(114) = -5.77, p < .001, d = -0.94$.

Although Hypothesis 1 was specifically about the increased usage of direct lines based on level of attractiveness, the hypothesis implies decreased usage in innocuous and flippant lines by attractive individuals compared to unattractive persons. Separate repeated-measures ANOVAs were used to explore differences in expected opening gambit usage between stimulus type (attractive, less attractive) and target stimulus gender (men, women; between-subject) for the innocuous and flippant opening gambits. For the innocuous lines, a main effect was observed for both stimulus type (attractive: $M = 9.64, SD = 2.01$; less attractive: $M = 11.75, SD = 4.55$), $F(1, 114) = 8.91, p = .003, \eta^2_p = .07$, and target stimulus gender (women: $M = 12.29, SD = 3.05$; men: $M = 10.11, SD = 1.83$), $F(1, 114) = 21.94, p < .001, \eta^2_p = .16$. The interaction between these variables was not significant, $F(1, 114) = 0.56, p = .46, \eta^2_p = .005$. These differences can be seen in the middle panel of Figures 2 and 3.

The 2 x 2 repeated-measures ANOVA exploring differences in expected usage for flippant lines across stimulus type and target stimulus gender revealed a main effect of stimulus type (attractive: $M = 4.62, SD = 2.25$; less attractive: $M = 7.28, SD = 4.35$), $F(1, 114) = 13.01, p < .001, \eta^2_p = .10$, target stimulus gender (women: $M = 5.23, SD = 2.33$; men: $M = 6.22, SD = 2.16$), $F(1, 114) = 4.60, p = .03, \eta^2_p = .04$, and the interaction between these two variables, $F(1, 114) = 8.34, p = .005, \eta^2_p = .07$.

Similar to the analyses for direct opening gambits, a series of $t$ tests explored the interaction between attractiveness and target stimulus gender. A dependent-measures $t$ test found no difference in mean usage of flippant lines between attractive ($M = 5.03, SD = 2.21$) and unattractive ($M = 5.42, SD = 4.12$) women, $t(30) = .46, p = .65, d = 0.09$. The mean difference between attractive ($M = 4.47, SD = 2.26$) and unattractive ($M = 7.96, SD = 4.26$) men was statistically significant, $t(84) = 6.11, p < .001, d = -0.68$.

A pair of independent-samples $t$ tests explored the difference in expected flippant gambit usage
between men and women at each level of attractiveness. No difference was found in the expected usage of flippant gambits for attractive women (M = 5.03, SD = 2.21) or men (M = 4.47, SD = 2.26), \( t(114) = 1.19, p = .24, d = .25 \). However, a significant difference was observed in the expected usage of flippant lines by less attractive women (M = 5.42, SD = 4.12) compared to men (M = 7.96, SD = 4.26), \( t(114) = -2.87, p = .005, d = .61 \). See the far left columns of Figure 3 for the expected differences in usage of flippant lines by less attractive and attractive women and men. Overall, the analysis of this interaction suggests that unattractive men are expected to use flippant opening gambits more than women, regardless of attractiveness level, or attractive men.

**Hypothesis 2**
The second hypothesis proposed that less attractive individuals would be predicted to present an innocuous or flippant opening gambit significantly more than a direct opening gambit. This hypothesis was evaluated using a repeated-measures ANOVA with gambit type (3 levels) as the independent variable and expected number of times each type of gambit was expected to be used as the dependent variable. This test only used the ratings for the less attractive stimuli in order to be consistent with the hypothesis. The main effect of gambit type was significant, \( F(2, 114) = 51.88, p < .001, \eta_p^2 = .48 \). The mean number of times each of the three opening gambits were selected was: flippant: \( M = 7.28, SD = 4.35 \); innocuous: \( M = 11.75, SD = 4.55 \); and direct: \( M = 5.94, SD = 2.87 \). Follow-up dependent measures \( t \) tests found significant differences between all pairings of the three gambit types: funny–innocuous, \( t(115) = 5.75, p < .001, d = .53 \); funny–direct, \( t(115) = 2.49, p = .14, d = .24 \); innocuous–direct, \( t(115) = 10.01, p < .001, d = .94 \).

Although Hypothesis 2 focused specifically on the expectations for the less attractive stimuli, we analyzed differences in the expectations for the three opening gambit types for the attractive stimuli as well. Extending Hypothesis 2, we expected that attractive individuals would use direct opening gambits more often than flippant or innocuous. The mean expected usage for the three opening gambits for the more attractive stimuli were: flippant: \( M = 4.62, SD = 2.25 \); innocuous: \( M = 9.64, SD = 3.01 \); and direct: \( M = 10.84, SD = 3.33 \). These differences were statistically significant, \( F(2, 114) = 161.73, p < .001, \eta_p^2 = .74 \). Follow-up pairwise testing with dependent-measures \( t \) tests found significant differences between the means for all three gambit types: funny–innocuous, \( t(115) = -12.72, p < .001, d = -1.21 \); funny–direct, \( t(115) = -4.27, p < .001, d = -1.36 \); innocuous–direct, \( t(115) = -2.17, p = .03, d = -0.20 \).

**Discussion**
Expectations often shape how people respond to different social interactions. Because attractiveness is one of the first characteristics people notice in others, expectations tied to physical attractiveness can be some of the most important in shaping social behavior. The purpose of this project was to investigate the effects of perceived attractiveness on opening gambit expectations. Results from the study showed that attractive people, overall, were expected to use direct opening gambits (Hypothesis 1; top panel of Figure 2). People rated as less attractive were expected to employ innocuous opening gambits, with some support for flippant lines (Hypothesis 2; middle and bottom panels of Figure 2). However, these expectations were dependent
on the gender of the stimulus target. Attractive men were expected to use direct gambits more often than unattractive men, but the difference in direct opening gambit usage did not differ across the pictured attractive and unattractive women. When the sample stimulus was an attractive woman, participants expected the model to use innocuous opening gambits most often. When the sample stimulus was a man, however, participants selected the direct opening gambit as the most expected opening gambit. This interaction was different for the less attractive stimuli. Less attractive men were expected to use flippant lines more often than less attractive women, but innocuous lines were selected most often for both sexes. Although flippant opening gambits are the stereotype for pickup lines, these opening gambit types were the least expected for three of the four groups. The exception being that less attractive pictures of men were expected to be paired with flippant lines more than direct lines.

The results from our study give researchers insight on the intuitive nature of humans’ perceptions of others in a social context. Based on characteristics of the initiator, participants had specific expectations. Attractive men were expected to use direct opening gambits, but attractive women were expected to use innocuous lines (later in this discussion, we discuss the importance of exploring the very likely impact of local culture on these expectations). People tend to be most comfortable when their expectations for social situations are met. This research clarified what the expectations are for people’s initiation of social interactions based on how their level of attractiveness is judged. The findings have implications for understanding social stereotyping and discrimination factors among across groups and individuals. An important consideration of the current study is the application of the Expectancy Violations Theory (EVT). The EVT is an interpersonal communication theory that suggests positive expectancy violations can result in more favorable outcomes than expectancy confirmations (Burgoon, 2016). In the context of the current research, the EVT theory might suggest that less attractive individuals using direct opening gambits would be a violation of the social expectation, which may result in improved social outcomes.

Opening gambit expectations may similarly lead to self-fulfilling prophecies in how people respond to and initiate social interactions (Downey, Freitas, Michaelis, & Khouri, 1998). When approached by someone that is perceived to be attractive, the individual being approached may unconsciously signal more openness to the social interaction (orienting toward the individual, increased eye contact). The initiator would then perceive less threat in the social interaction, responding with a more direct opening line. Thus, the initiator expected a direct approach from an attractive individual and behaved in a way that engendered this type of behavior. At the same time, a person who self-perceives as less attractive may nervously approach the social interaction, creating discomfort in the recipient, and thus resulting in unintended withdrawal signals (e.g., orienting away from the participant, decreased eye contact). In both of these examples, the expectations of the initiator and the recipient influence the social interaction, even before verbal exchange begins.

Due to the possibility that the observed expectancies may be involved in interpersonal self-fulfilling prophecies or expectancy violations, a particularly important line of future research should explore the impact of violating the expectation based on facial attractiveness. For example, what is the impact of a less attractive individual using a direct opening gambit (contrary to expectations) or an attractive person using a flippant or innocuous gambit? Would a lack of congruence between expectation and behavior have a detrimental effect on the development of the relationship? Although the goals of the current study were to describe the interaction between facial attractiveness and opening gambit expectations, future research needs to directly explore questions about how these expectations develop, how they might impact social interactions, and how violations of these expectations impact the social dynamics.

Future research should also examine other traits that might influence the opening gambit expectations. Candidate physical traits may be shoulder-to-waist ratios for men, and waist-to-hip ratios for women (Kościnski, 2014). Nonphysical factors that may impact the expectations of opening gambits include personality factors or economic status, thus future research would benefit from the investigation of these variables as well. Additionally, racial and ethnic background of the models and participants could be important variables. Although models from different racial backgrounds were used, these data were not analyzed along this dimension. No racial information was collected for participants. Future studies should include these variables, as well as seek a diverse sampling population because cultural context is likely an influence on opening gambit expectations and
Attractiveness and Opening Gambits | Wood, Charlton, Goodman, and Thompson

receptivity. As mentioned in the study, social context plays an important role in the selection of opening gambits, and future research should examine the association between the different styles of gambits and different environments and social situations. There were several limitations to this study. First, we excluded participants who reported an equal stimulus sexual attraction to men and women. These participants were excluded because it was unclear which set of stimuli they should be shown. Future research should include these participants, providing either a mixed set of stimuli or asking them to identify five attractive or five less attractive stimuli (independent of sex) for use in the study. Similarly, recruiting a sample that includes more participants with nonheterosexual preferences would allow for an exploration of differences in expectations across sexual preference categories. Finally, we did not ask participants for information regarding their racial or ethnic background. It is possible that expectations differ across these categories, and this should be explored in future work.

This study provided an important unexplored component to current literature in the field of social psychology by demonstrating a critical correlation between facial attractiveness and social stereotypes. In addition, the findings of the current study gave further insight to the mate-selection process and the impressive ability to form intuitive expectations of others’ manifested behaviors based on their perceived ratings of facial attractiveness. These expectations in the social context can control how recipients perceive initiators based on the initiators’ social attractiveness. These perceptions may then influence recipients’ willingness to engage with these individuals. This project created a descriptive foundation for additional research into the psychological factors involved in initiating social interactions as well as new insights into how attractiveness influences social expectations.

References
Wade, T. J., Butrie, K. L., & Hoffman, M. K. (2009). Women’s direct opening lines are perceived as most effective for an affair and individual differences, 2, 145–149. https://doi.org/10.1037/a0016578

The authors report no conflict of interest in the completion of this project.

This project was completed as a capstone project in PSYC 3340: Research Methods Laboratory. A special thanks to our classmates for their assistance in designing the materials, pilot testing the studies, and reviewing the manuscript. Also, our sincere appreciation toward Sarah Lindeman for comments on multiple drafts of this manuscript.

Correspondence concerning this article should be addressed to Shawn R. Charlton, University of Central Arkansas, Department of Psychology and Counseling, 201 Donaghey Ave, Conway, AR 72035. E-mail: scharlton@uca.edu

APPENDIX

Types of Opening Gambit Styles

<table>
<thead>
<tr>
<th>Opening Gambit Type:</th>
<th>D = Direct; I = Innocuous; and F = Flippant</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Group 1:</strong></td>
<td></td>
</tr>
<tr>
<td>What do you do for a living? (I)</td>
<td></td>
</tr>
<tr>
<td>Can I sit with you so I won’t get hit on? (F)</td>
<td></td>
</tr>
<tr>
<td>Hey beautiful, how are you today, would you like to go out for a cup of coffee? (D)</td>
<td></td>
</tr>
<tr>
<td><strong>Group 2:</strong></td>
<td></td>
</tr>
<tr>
<td>Is anyone sitting here? (I)</td>
<td></td>
</tr>
<tr>
<td>If you were a triangle you’d be acute one. (F)</td>
<td></td>
</tr>
<tr>
<td>Can I buy you a drink? (D)</td>
<td></td>
</tr>
<tr>
<td><strong>Group 3:</strong></td>
<td></td>
</tr>
<tr>
<td>Don’t I know you from somewhere? (I)</td>
<td></td>
</tr>
<tr>
<td>Do you know how much a polar bear weighs? Enough to break the ice, Hi I’m (insert name here). (F)</td>
<td></td>
</tr>
<tr>
<td>I think I follow you on [social media platform]. We should get to know each other. (D)</td>
<td></td>
</tr>
<tr>
<td><strong>Group 4:</strong></td>
<td></td>
</tr>
<tr>
<td>Hey I’m new around here, where’s your favorite place? (I)</td>
<td></td>
</tr>
<tr>
<td>You breathe oxygen too? We have so much in common. (F)</td>
<td></td>
</tr>
<tr>
<td>I just have to tell you, you have amazing eyes. (D)</td>
<td></td>
</tr>
<tr>
<td><strong>Group 5:</strong></td>
<td></td>
</tr>
<tr>
<td>Are you good at _____? I need a study partner and I think we could help each other out. (I)</td>
<td></td>
</tr>
<tr>
<td>Is your name Ariel? Because I think we mermaid for each other. (F)</td>
<td></td>
</tr>
<tr>
<td>I saw you and I just had to come talk to you. (D)</td>
<td></td>
</tr>
</tbody>
</table>
Find your career.

Eight graduate degree programs and four certificates in Educational Psychology

**PhD in Educational Psychology**
Engage in the science of learning. Prepare for a career where you can use your knowledge of human learning and development to help shape the school environment and public policy. Core program areas include learning, motivation, and research design.

**MS or MA in Educational Psychology**
Broaden your ability to apply psychological principles to a variety of professional contexts or prepare for your future doctorate in social science.

**MS in Quantitative Psychology**
Do you like numbers, statistics, and social science? Prepare for a career in research, assessment, and data analysis. Develop proficiency in advanced statistical techniques, measurement theory, and data analytics.

**PhD in School Psychology** (five-year program)
Prepare for a career as a licensed psychologist. Gain competencies in health service psychology to work in schools, private practice, or hospital settings. Accredited by the American Psychological Association (APA)** and approved by the National Association of School Psychologists (NASP). Scientist-practitioner model with advocacy elements. Specializations available.

**MA/EdS in School Psychology** (three-year program)
Be immersed in community engaged, real-world field experiences and intervention opportunities in our scientist-practitioner-advocate program. Leads to licensure as a school psychologist. Approved by NASP and the National Council for Accreditation of Teacher Education (NCATE).

**MA in School Counseling** (two-year program)
Be a leader and advocate for educational equity for all students in PK-12 schools. Leads to licensure as a school counselor. Accredited by the Council for Accreditation of Counseling and Related Educational Programs (CACREP) and nationally recognized by The Education Trust as a Transforming School Counseling program.

**Certificates**
- High Ability/Gifted Studies,*
- Human Development and Learning,*
- Identity and Leadership Development for Counselors,*
- Neuropsychology*

Graduate assistantships and tuition waivers are available.

bsu.edu/edpsy

*Online programs are available.

**Questions related to the PhD in school psychology's accreditation status should be directed to the Office of Program Consultation and Accreditation, American Psychological Association, 750 First St. NE, Washington, D.C. 20002; (202) 336-5979; apaaccred@apa.org; or apa.org/ed/accreditation.
"MY JOB IS NOT JUST TO TEACH, BUT ALSO TO HELP STUDENTS SEE THEIR INNER STRENGTHS."

At the College of Clinical Psychology at Argosy University, we believe in a practitioner-scholar model of training. Our programs offer a rigorous curriculum grounded in theory and research, while also offering real-world experience. What’s more, all our PsyD programs have received accreditation from the American Psychological Association (APA), certifying that they meet the industry’s standards.

Learn more at clinical.argosy.edu/psichi

Arizona School of Professional Psychology at Argosy University
American School of Professional Psychology at Argosy University | Southern California
American School of Professional Psychology at Argosy University | San Francisco Bay Area
Florida School of Professional Psychology at Argosy University
Georgia School of Professional Psychology at Argosy University
Hawaii School of Professional Psychology at Argosy University
Illinois School of Professional Psychology at Argosy University | Chicago
Illinois School of Professional Psychology at Argosy University | Schaumburg
Minnesota School of Professional Psychology at Argosy University
American School of Professional Psychology at Argosy University | Northern Virginia

DR. NAHID AZIZ
Associate Professor at the American School of Professional Psychology at Argosy University | Northern Virginia

Dr. Aziz is committed to mentorship, training, and addressing issues relevant to the ethnic and racial diversity.

*The Doctor of Psychology in Clinical Psychology Program at Argosy University, Atlanta, Chicago, Hawaii, Orange County, Phoenix, San Francisco Bay Area, Schaumburg, Tampa, Twin Cities and Northern Virginia is accredited by the Commission on Accreditation of the American Psychological Association (APA). Questions related to the program’s accredited status should be directed to the Commission on Accreditation, Office of Program Consultation and Accreditation, American Psychological Association, 750 First Street, NE, Washington DC 20002-4242. Phone: (202) 336-5579. E-mail: apacommission@apa.org / Web: www.apa.org/ed/accreditation

Argosy University is accredited by the WASC Senior College and University Commission (855AltaSt.,Suite105,Artsmeta,CA94601,wsac.org). Programs, credential levels, technology, and scheduling options are subject to change. Not all online programs are available to residents of all U.S. States. Administrative office: Argosy University, 601 South Lewis Street, Orange, CA 92868 ©2018 Argosy University. All rights reserved. Our email address is materialsreview@argosy.edu
See and Share  
*Psi Chi Journal’s First-Ever Special Issue*

www.psichi.org/page/232JNSpecialIssue18

Learn about “replication crisis” and how earning Open Science Badges can reinforce your research practices. Led by Special Invited Editor Dr. Steven V. Rouse, this unique, free issue features a variety of articles awarded with badges for sharing Open Data, Open Materials, Preregistering their research, and conducting Replication studies.

Authors of all future *Psi Chi Journal* articles are encouraged to participate in our ongoing badges program, which was created by members of the Open Science Collaboration. Any badges received will be prominently displayed on the first page of published *Psi Chi Journal* articles, as shown throughout the new special issue.

Learn more about the badges at www.psichi.org/page/journal_Badges

View *Psi Chi Journal* submission guidelines at www.psichi.org/?page=JN_Submissions

---

**Special Issue Call for Abstracts:**  
**Education, Research, and Practice for a Diverse World**

Today, consider submitting an abstract for a potential empirical manuscript to be published in *Psi Chi Journal’s* next special issue. This special issue welcomes articles that examine the effects and advancement of diversity and inclusion in classrooms, educational materials, and technology.

- Abstracts are due **September 15**.
- Selected manuscripts are due **October 15**.
- Publication for the issue is anticipated for **Summer 2019**.

A vast amount of psychological research has indicated the importance of embracing diversity-affirming stances with regard to racial and ethnic diversity, sexual and gender diversity, religious diversity, disability status, SES and social class, refugee and immigrant status, among many others. This special issue is in direct support of the new 2018–19 Psi Chi theme of “Leveraging Technology to Advance Diversity and Inclusion.”

We look forward to reviewing your abstracts! Submit your abstracts to Debi.Brannan@psichi.org.

---

*Psi Chi Journal of Psychological Research*

COPYRIGHT 2018 BY PSI CHI, THE INTERNATIONAL HONOR SOCIETY IN PSYCHOLOGY (VOL. 23, NO. 3/ISSN 2325-7342)
PSI CHI Advertising Contract: Psi Chi Journal

OUR JOURNAL
Advertising in Psi Chi Journal allows you to connect with established psychology researchers and mentors, as well as undergraduates and graduate students striving to build a career in one of the many areas of research. People regularly visit our journal online to:
• view current and past issues,
• submit their research for publication,
• learn about reviewing for Psi Chi Journal, and
• share invited editorials as teaching tools in the classroom.

All issues and advertisements are permanently free online to both members and nonmembers alike. During the 2017–18 fiscal year, pschi.org received almost 1.5 million page views, ensuring that high-achieving students and professionals will see your content for years to come.

To further enhance the visibility of our journal, latest issues and calls for submissions are regularly featured in Psi Chi Digest e-mails (177,000+ subscribers) and on our four social media platforms:
• Facebook (22,500+ followers)
• Twitter (4,700+ followers)
• LinkedIn (10,200+ followers)
• Instagram (1,000+ followers)

Articles are also indexed in PsycINFO, EBSCO, and CrossRef databases—essential tools that researchers use to search for millions of psychology-related articles. This makes Psi Chi Journal a key place to communicate your message with our Professional Organization’s three quarters of a million lifetime members and far beyond.

AD SPECIFICATIONS
Digital format: PDF, EPS, and TIF
Resolution: 300 dpi | B&W line art—1,200 dpi
Black & white ads (no RGB or 4-color process)
PDF settings: Press quality, embed all fonts

DEADLINE/BILLING
Payment due upon receipt of invoice.

CONTACT
Submit contract by e-mail to
Susan Iles
Advertising Sales
Psi Chi Central Office
E-mail: susan.iles@pschi.org
Phone: 423-771-9964
Publish Your Research in *Psi Chi Journal*
Undergraduate, graduate, and faculty submissions are welcome year round. Only the first author is required to be a Psi Chi member. All submissions are free. Reasons to submit include

- a unique, doctoral-level, peer-review process
- indexing in PsycINFO, EBSCO, and Crossref databases
- free access of all articles at psichi.org
- our efficient online submissions portal

View Submission Guidelines and submit your research at [www.psichi.org/?page=JN_Submissions](http://www.psichi.org/?page=JN_Submissions)

---

**Become a Journal Reviewer**
Doctoral-level faculty in psychology and related fields who are passionate about educating others on conducting and reporting quality empirical research are invited become reviewers for *Psi Chi Journal*. Our editorial team is uniquely dedicated to mentorship and promoting professional development of our authors—Please join us!

To become a reviewer, visit [www.psichi.org/page/JN_BecomeAReviewer](http://www.psichi.org/page/JN_BecomeAReviewer)

---

**Resources for Student Research**
Looking for solid examples of student manuscripts and educational editorials about conducting psychological research? Download as many free articles to share in your classrooms as you would like.

Search past issues, or articles by subject area or author at [www.psichi.org/?journal_past](http://www.psichi.org/?journal_past)

---

**Add Our Journal to Your Library**
Ask your librarian to store *Psi Chi Journal* issues in a database at your local institution. Librarians may also e-mail to request notifications when new issues are released.

Contact PsiChiJournal@psichi.org for more information.

---

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