Evaluating Television Shows: The Influences of Commercials and Customized Ratings on Perceived Enjoyment

A dissertation presented to
the faculty of
the Scripps College of Communication of Ohio University

In partial fulfillment
of the requirements for the degree
Doctor of Philosophy

Jeremy Saks
April 2017
© 2017 Jeremy Saks. All Rights Reserved.
This dissertation titled
Evaluating Television Shows: The Influences of Commercials and Customized Ratings on Perceived Enjoyment

by
JEREMY SAKS

has been approved for
the E. W. Scripps School of Journalism
and the Scripps College of Communication by

Carson B Wagner
Assistant Professor of Journalism

Scott Titsworth
Dean, Scripps College of Communication
ABSTRACT

SAKS, JEREMY, Ph.D., April 2017, Journalism

Evaluating Television Shows: The Influences of Commercials and Customized Ratings on Perceived Enjoyment

Director of Dissertation: Carson B Wagner

This dissertation, comprised of two experimental studies, focuses on the impact of pairing targeted advertisements and ratings with television programming for college-aged individuals in terms of media enjoyment. The first experiment, building off of Saks (2013), examines how pairing television shows with either age-congruent or age-incongruent advertisements can influence participants’ enjoyment of shows from three different genres outside of the participants’ age demographic. The second experiment moves into the realm of targeted ratings, similar to those utilized by websites like Netflix. That experiment explores how individuals are influenced by manipulated ratings telling them how much they will supposedly like a show. The results from the first study, which utilized Mandler’s Discrepancy/Evaluation theory, show no statistical significance in terms of change in enjoyment when viewers’ saw shows outside of their age demographic paired with either age-congruent or age-incongruent advertisements. The second study, however, showed statistical significance for individuals’ enjoyment being manipulated by randomly assigned supposedly targeted ratings. Various explanations for both phenomena are discussed.
DEDICATION

Dedicated to my family (which includes my wife-to-be, Krista),

my committee,

and my cats.

Thank you!

Without your support, I never could have accomplished what I have and what I will.
ACKNOWLEDGMENTS

There are so many people to thank and acknowledge that I am sure I will forget someone and not give others their due respects. However, I must acknowledge and thank my fiancée, Krista, who has been there for me and put up with my extended time in graduate school while giving emotional, mental, physical, and monetary support. The same goes for my parents who have been on my side throughout the process and pushing me to succeed while having read countless revisions and drafts of work throughout the years.

I must also thank my support at Ohio University. Foremost, my chair, Carson B Wagner, who has helped expand my mind immensely in such a way that I can only hope to one day pass that on to one of my own students. Having worked with me through many classes, my thesis, and now my dissertation, he has had an incredible impact on me and my academic journey. This dissertation would not have been possible without his laboratory and the E.W. Scripps Participant Pool that he ran. Similarly, Jatin Srivastava has been a constant source of inspiration, knowledge, and guidance throughout my time at Ohio University. Despite never having officially taken a course with Dr. Srivastava, he has been very open and helpful, allowing me to use his laboratory on multiple occasions while also serving on my thesis, comprehensive exam, and dissertation committees. He has been exceedingly nice and pleasant while helping, especially with methodological and statistical issues, and has done more than anyone else to create a solid analytic footing for which I can utilize in the future. Dr. Roger Cooper has provided great support, regardless of being outside of my department within the College
of Communication. Dr. Ronaldo Vigo has been very helpful for me to better understand perspectives outside of the communication paradigm.

I also want to thank Dr. Michael Sweeney as the graduate director and his tireless work ethic to help all of those in the graduate program. Within the journalism department, the administrative associates (Debbie Depeel, Sharon Nickels, Julie Conrad, and Rebecca Fox) and student workers help keep everything working behind the scenes.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstract</td>
<td>iii</td>
</tr>
<tr>
<td>Dedication</td>
<td>iv</td>
</tr>
<tr>
<td>Acknowledgments</td>
<td>v</td>
</tr>
<tr>
<td>List of Tables</td>
<td>ix</td>
</tr>
<tr>
<td>List of Figures</td>
<td>x</td>
</tr>
<tr>
<td>Introduction</td>
<td>1</td>
</tr>
<tr>
<td>Study 1: Advertising Content, Context, and Congruence with Age</td>
<td>5</td>
</tr>
<tr>
<td>Literature Review</td>
<td>5</td>
</tr>
<tr>
<td>Methodology</td>
<td>14</td>
</tr>
<tr>
<td>Results</td>
<td>18</td>
</tr>
<tr>
<td>Discussion</td>
<td>21</td>
</tr>
<tr>
<td>Study 2: Effects of Customized Ratings on User Evaluations of Television Shows</td>
<td>28</td>
</tr>
<tr>
<td>Literature Review</td>
<td>28</td>
</tr>
<tr>
<td>Methodology</td>
<td>33</td>
</tr>
<tr>
<td>Experimental Design</td>
<td>33</td>
</tr>
<tr>
<td>Sampling</td>
<td>34</td>
</tr>
<tr>
<td>Stimulus Design</td>
<td>34</td>
</tr>
<tr>
<td>Measurements / Instruments</td>
<td>36</td>
</tr>
<tr>
<td>Procedure</td>
<td>37</td>
</tr>
<tr>
<td>Results</td>
<td>38</td>
</tr>
<tr>
<td>Discussion</td>
<td>43</td>
</tr>
<tr>
<td>General Discussion</td>
<td>56</td>
</tr>
<tr>
<td>Theoretical Implications</td>
<td>56</td>
</tr>
<tr>
<td>Methodological Implications</td>
<td>58</td>
</tr>
<tr>
<td>Practical Implications</td>
<td>61</td>
</tr>
<tr>
<td>Research Limitations and Future Studies</td>
<td>62</td>
</tr>
<tr>
<td>Conclusion</td>
<td>65</td>
</tr>
<tr>
<td>References</td>
<td>66</td>
</tr>
<tr>
<td>Appendix A: Example Implicit Association Test for Study 1</td>
<td>84</td>
</tr>
<tr>
<td>Appendix B: Explicit Evaluation for Study 1</td>
<td>94</td>
</tr>
<tr>
<td>Table</td>
<td>Description</td>
</tr>
<tr>
<td>-------</td>
<td>-------------</td>
</tr>
<tr>
<td>Table 1</td>
<td>Skin Conductance Levels by Group</td>
</tr>
<tr>
<td>Table 2</td>
<td>Mean and Standard Deviation on Seven-Point Scale by Condition for Each Show</td>
</tr>
<tr>
<td>Table 3</td>
<td>Mean and Standard Deviation of Indexes by Condition for Each Show</td>
</tr>
<tr>
<td>Table 4</td>
<td>Frequency of Participant Responses Statement “I Watch (Program Name) Often”</td>
</tr>
</tbody>
</table>
LIST OF FIGURES

Figure 1. Mean Response to the Videos by Condition on Seven-Point Scale ..................45
Figure 2. Mean Response to the Videos by Condition on Eight-Item Indexed Scale ....46
Figure 3. Mean Responses to Affective and Behavioral Statements by Condition .......48
INTRODUCTION

Advertisements are media content as much as are the programs and other media in which they are embedded. Although users are likely more interested in the shows, social media, newsfeeds, etc. that comprise what media outlets might call “primary content,” the overall experience in viewing media includes both primary content and the “secondary content” of advertisements and presentational packaging. Research has been conducted about the motivational effects of advertising on its audience, examining purchase intent (Ohanian, 1991; Kamins & Marks, 1987; Park, Lennon, & Stoel, 2005), brand awareness (Percy & Rossiter, 1992; Hoyer & Brown, 1990; Keller, 1993), and other dependent variables ranging from awareness to choice. There is a wealth of research on numerous areas of advertising and motivation, but the contextual effects of advertisements’ presentation seem similarly worthy of rigorous, systematic scientific research. The surrounding content – which, from such a vantage point, could be called “secondary” – can affect the ways in which ads are perceived and received. That generally would occur due to different types of “priming” (see, e.g., Berkowitz, 1984; Kahneman & Tversky, 1984; Iyengar, Peters, & Kinder, 1982) or rather, media content that precedes and thereby can influence the ways in which advertising affects those same dependent variables.

The past research on media and advertising, particularly with “priming” in mind, has led to many questions, some examined and some not. For instance, we might ask: 1) “Does a program’s ‘primary content’ influence the viewers' perception of an ad's efficacy?”; 2) “In what ways and regarding which independent variables?”; 3) “Does choice of advertisements affect how a viewer perceives the primary content?”; 4) “Does
the content’s genre modify any type of effect experienced from secondary content?”; or 5) “Does the fact that a particular presentation is *recommended* for a viewer affect how much the viewer likes it?” That last question focuses on how recommendation agents, which have become increasingly prevalent on the web, may impact how viewers evaluate programming instead of simply suggesting content.

Most prior research on context has focused on the first two of these questions: the effect that content has on how viewers evaluate certain ads (Aylesworth & MacKenzie, 1998; Coulter, 1998; Zanjani, Diamond, & Chan, 2011; Dahlén, Rosengren, Törn, & Öhman, 2008; Yoon, 2013; Alden, Mukherjee, & Hoyer, 2000; Goldberg & Gorn, 1987; Gunter, Furnham, & Beeson, 1997). An exhaustive search of the literature yielded no prior research on the third question except Saks (2013), which examined the effects of advertisement viewing on contextual content. The study showed that audiences tended to evaluate more favorably those television shows that paired content with advertisements congruent with the program’s demographics (more specifically, age group) than those paired with demographically incongruent ads.

This dissertation is comprised of two controlled, laboratory experiments that extend prior research in various ways – promoting the strength of exploratory experimentation by demonstrating that given effects *can* happen, given (a) sufficiently strong stimuli and (b) necessarily sensitive dependent variable measures (Sundar, 1999). The first experiment furthers previous research by reversing the types of (in)congruity in Saks’ (2013) work. By showing a television program *outside* of a given set of viewers’ demographic but including advertisements that either match the *targeted age group of the*
show \((incongruent\) with participants) or within the age group of the participants \((congruent\) with participants), the experiment will expand upon the potential influence of pairing varying advertisements with television content. Additionally, the experiment will add the additional independent variable of content’s genre to better understand different types of media enjoyment that can occur across genres.

A second experiment introduced an element of personalization not available through demographic generalization. The manipulation, instead, included telling participants that they received a “customized experience” based on their past media usage. As such, the potential effects of perceived more accurate form of targeting were measured. The first experiment examined whether having a show incongruent with the viewers’ demographic would be viewed more positively with ads aimed at that show’s target demographic or with those of the viewers’ demographic. The second experiment moved beyond standard demographic targeting into the realm of “personalization.”

Saks’ (2013) initial ad (in)congruity study was a step toward better understanding the effects of targeting, but ultimately that and Study 1 of the dissertation will, together, pose more questions for future research. Study 2 examines how other forms of information – specifically, recommended content ratings – can influence how individuals view related content by building expectations within the users.

Study 2 provided users a manipulated rating prior to viewing a show to see if the suggestion of how the user will rate the content ultimately modifies the user’s reported perception of said show. Many media displayed on the Internet include ratings relating to how well the communicators believe users will like the target content, based on the
evaluations by other users and the utilization of sophisticated algorithms. For example, Netflix has users rate previously viewed content on a five-point scale so that the site can give predictions (on that same five-point scale) about how the service believes the viewer will enjoy other available content. Such ratings are an important part of popular media websites including Netflix, Hulu, and Amazon Prime Instant Video. In 2006 Netflix launched a widely-publicized competition that promised a million-dollar grand prize for an algorithm that could out-perform Netflix's own (Bennett & Lanning, 2007). In the same ways that producers, distributors, and users should have an interest in understanding how advertisements may alter perceptions of content, those same parties would be interested in knowing how ratings and suggestions can influence users (Saks, 2013). Together, these two experiments will help to better understand the outside influence on perceptions of media content in ways that commonly occur in today’s media environment.
STUDY 1: ADVERTISING CONTENT, CONTEXT, AND CONGRUENCE WITH AGE

Literature Review

Saks (2013) used Mandler’s (1982; 1984; 1989; 1990; 1992; 1994) research about his affective theories of discrepancy and evaluation to show that participants evaluated a television show significantly differently when that show was paired with age-congruent and age-incongruent advertisements. Discrepancy/Evaluation Theory of Emotion states that when experiencing (a) stimulus(i) different from what is expected, individuals will have an arousal response that is modified by an evaluation based on prior experience in forming an emotional response (Weiner, 1986). In the context of Study 1, individuals should have expectations as to what type of commercials often appear with specific types of content. Those expectations will have been formed into mental schemas (Markus, 1977) through previous exposures to television. Prior research using Mandler’s (1982) theory has utilized expectations created within the laboratory prior to the measurement (MacDowell & Mandler, 1989; Gaver & Mandler, 1987).

The foundation of Mandler’s (1990) Discrepancy/Evaluation of Emotion can be broken down into four parts. First, participants must have an expectation of what will occur in a given scenario based on previously created mental schema(s) (Markus, 1977). Second, a discrepancy from that expectation leads to a physiological arousal response. Third, the participant’s evaluations modify the physiological response to determine direction of emotional response, as is the case with attributional theories of emotion (see, e.g., Weiner, 1986). Finally, the emotional response modifies the participant’s perception
of the subject material. Therefore, in regard to television shows and commercials, Mandler’s (1990) theory that discrepancies will result in positive or negative evaluations, depending on the extent to which the new stimulus deviates from one’s related schema, should lead to alterations within the participants’ responses. The theory has been used in multiple studies on different types of media content, including video games (MacDowell & Mandler, 1989), music (Gaver & Mandler, 1987), website banner advertisements (Moore, Stammerjohan, & Coulter, 2005), and print advertisements (Stafford, Walker, & Blasko, 1996).

Despite using “media enjoyment” as the dependent measure in multiple studies over the years, mass media researchers have not come to understand the phenomena of enjoyment well enough to create a uniform conceptualization of it (Nabi & Krcmar, 2004). Various theories have been posited in recent years (e.g., Green, Brock, & Kaufman, 2004; Sherry, 2004; Nabi & Krcmar, 2004; Tamborini, Bowman, Eden, Grizzard, & Organ, 2010; Denham, 2004; Vorderer, Klimmt, & Ritterfeld, 2004). Nabi and Krcmar (2004) offer the most complete understanding of the concept by breaking “enjoyment” down into three components based on those which have been considered the three main functions performed and directed by the human brain since the philosophy of Plato: affective, cognitive, and behavioral (Allport, 1968; Hilgard, 1980). Those functions have been ported to psychology and persuasion by McGuire (1968) and, more specifically, to advertising by Leckenby and Wedding (1982). Many studies have focused solely on the affective dimensions of enjoyment (Nabi & Krcmar, 2004) and of
those studies many have used only one or two questions to measure the dependent variable (Weaver & Wilson, 2009; Weaver, 2011).

For Study 1, it is possible to isolate the affective and cognitive components and offer insight into future behaviors through participants’ responses as well as implicit attitude measurements by determining how age-congruent advertisements affect users' perceptions of a program. Considering the literature about attitudes and subtle persuasive content stimuli that may be consciously imperceptible (Wagner, 2001), it is important to include both an implicit and explicit measure in the study to better understand the different types of enjoyment that the participants may experience even if they cannot express them explicitly (see, e.g., Nisbett & Wilson, 1977). Additionally, due to its theoretical reliance on arousal response, it is necessary to include skin conductance level, which measures physiological arousal (see, e.g., Sundar & Wagner, 2002), to demonstrate Mandler’s (1990) theory, most specifically.

There are a variety of similar theories that are important to consider when creating this study and analyzing the results. Lang’s (2006) limited capacity model of motivated mediated message processing (LC4MP) offers insight into the ways in which individuals process media messages based on the amount of cognitive resources that they have available, although so far the majority (if not all) of the research utilizing that theory deals with only one stimulus. The theory of processing fluency (Winkielman, Schwarz, Fazendeiro, & Reber, 2003) states that individuals will prefer stimuli that are easier to process and will be relatively averse to stimuli that are harder to process, which stems
from humans’ tendency to be “cognitive misers” (Fiske & Taylor, 1991) who have to expend more energy when they may not want to do so.

Individuals may (sub)consciously view the congruency (or lack thereof) with different processing styles, resulting in dissimilar emotions. Similarly, mood-as-information theory (otherwise known as “affect-as-information” theory and “feelings-as-information” theory; Schwarz & Clore, 1983) postulates that individuals will use their subconscious moods to inform how to process given information. Positive affect will lead to a less strict processing of information, while negative affect will signal an issue that may need to be changed, resulting in stricter information processing. Additionally, Expectancy-Disconfirmation theory (Cardozo, 1965; Anderson, 1973; Oliver, 1980) focuses on consumer satisfaction based on expectations that one brings to an experience.

Depending on the valence of their expectations, individuals may either be satisfied or dissatisfied if their expectations are not met (or disconfirmed). In the case of Study 1, individuals will have expectations about how a media package should be constructed based on past experiences with television and advertisements; however, those expectations may either be confirmed or disconfirmed based on the commercials included in the media presentation. All of these theories have numerous studies backing up their validity and may help extend the results from this study.

Saks (2013), the progenitor of the current proposed study, found that individuals exposed to age-incongruent advertisements paired with content targeted at college students led to individuals having lower explicit opinions of the content than those who saw age-congruent ads. In that study, the author examined how age-based demographic
(in)congruity between a television show and advertisements motivated individuals to think differently about the show. The television show stimulus in that study targeted an 18 – 25 age group. The two experimental conditions paired content with either advertisements for the show’s target age range (18 – 25) or commercials for senior citizens (aged 65 and above). The results showed age-based incongruity lowered the participants’ opinions of the show.

Although the results showed those who saw the ads for older adults liked the show less, it may not have been the content of the ads that changed their thinking. The results may have shown a propensity for individuals to like shows more when the advertisements and show are demographically congruent as a single package instead of contextually congruent for the viewers’ demographic.

Interestingly, the age-incongruent ads, aimed at senior citizens, elicited almost the opposite response from participants when using implicit measures, namely the Implicit Associations Test (Greenwald, McGhee, & Schwartz, 1998).

The results are qualified as “almost the opposite response” in that the results were not statistically significant, but did approach significance ($p = .092$). However there was no significant difference in the physiological responses measured, which may explain the lack of significance found in the implicit measures. Fazio and Towles-Schwen (1999) say that implicit association tests (IATs; Greenwald, McGhee, & Schwartz, 1998) measure affective differences, and Mandler's (1990) theory explains those differences as a function of differences in level of arousal in response to discrepancies. Because the physiological responses showed no significant difference in level of arousal, these
theories taken together would indicate that the lack of significance in implicit measures is not surprising. Mandler’s (1990) theory explains affective differences – which Fazio and Towles-Schwen (1999) say that implicit association tests (IATs; Greenwald, McGhee, & Schwartz, 1998) measure – as a function of differences in level of arousal, in response to discrepancies.

In considering Saks (2013), it is evident that the prior study was inconclusive as to the factors that contributed to the viewers’ responses. The purpose of Study 1 is to isolate those factors more fully. While the prior study showed participants a program targeted to their age group, Study 1, while using a similar college-aged audience, will use a program aimed at senior citizens. By examining the impact of a demographically incongruent television show, it is possible to rule out potential answers to additional questions raised by Saks (2013).

Those questions include: 1) “Will the effect still occur when the target show is demographically incongruent with the audience versus congruent?”; 2) “Did the participants prefer that there was a demographically congruent advertisements match with the content of the show as opposed to the mismatch between the incongruent advertisements with the content?”; and 3) “Will the difference in explicit vs. implicit measures persist under further experimentation?” It is difficult to know how participants will react in Study 1, as it is unclear whether they were influenced by the mismatch of congruence between the content and advertisements, incongruence between the advertisements and the demographic identity of the viewers, both, or something entirely different.
Mandler's (1990) Discrepancy/Evaluation of Emotion theory would indicate that the viewers, who have a previous mental schema (Markus, 1977) that advertisements are typically demographically congruent with programming content, would encounter a discrepancy when confronted with incongruent advertisements. That discrepancy should result in an arousal response that can be measured in Study 1. As such, Study 1 should provide additional evidence for understanding the phenomena with regard to pairing television content with differing commercials by examining a similar set of commercials, for a sample with similar demographics, while changing the content. To test these possibilities, based on the theory, the two main hypotheses for this study are:

H₁: Participants will explicitly evaluate a television show targeted at senior citizens differently when viewing that program with advertisements targeted at senior citizens as compared to viewing it with advertisements targeted at college-aged individuals.

H₂: Participants will implicitly evaluate a television show targeted at senior citizens differently when viewing that program with advertisements targeted at senior citizens as compared to viewing it with advertisements targeted at college-aged individuals.

One element that was not considered in Saks (2013) was variations in potential effects based on the influence of the shows’ genres. Genre, especially within the realm of television, is a difficult concept to define for a variety of reasons. (For an extensive discussion of the specific issues relating to television genre and a proposed cultural view of television genres, see Mittell, 2001). In particular, genres have been delineated from numerous sources depending on the narrative, the historical significance, the iconography and semiotics, and characters, among other factors (Neale, 2001). Additionally, with a multitude of different voices from the television industry that influence which genre is assigned to a television show (including the producer, director, actor, network executive,
network programmer, as well as others) it is difficult to have a singular genre without hybridization and shared elements (Turner, 2001). Given the various ways that genres can overlap, some scholars (Waller, 1988; Mittell, 2001; Mittell, 2004) have argued that genres should be examined from critical and cultural perspectives, as well as for textual connectedness and commonalities. Despite these definitional issues, a multitude of publications (Mills, 2004; Shim & Paul, 2007; Cohen & Weimann, 2000; Mittell, 2003; Savorelli, 2010) and reference books (Calvert, Casey, Casey, French, & Lewis, 2007; Rose, 1985; Creeber, 2015) have either outlined common elements shared by numerous genres and subgenres or used those genres as independent variables.

Various studies have analyzed viewers’ choices and usages in genres, often utilizing a uses and gratifications perspective and/or individual personality factors in the analyses (Weaver, 1991; Weaver, 2003; Hall, 2005; Hawkins et al., 2001; Shim & Paul, 2007).

From a media effects standpoint, however, not many studies have analyzed differences in genres. Potential difficulty in such studies comes from the diverse classifications of each individual genre and how the notable dissimilarities are tricky to control. Even within a specific genre, there are many subgenres that could have their own unique qualities that could shift the effects on the audience. That being said, some studies have attempted to better understand the effects that watching particular genres (or subgenres) have on individuals, most notably within the framework of cultivation theory, originally posited by George Gerbner (1969) and later formalized by Gerbner and Gross (1976). Those studies include examinations of reality television (Nabi, Biely, Morgan, &
Stitt, 2003; Egbert & Belcher, 2012), the cultivation of violent responses from participants after viewing dramatic television programming from multiple subgenres (Gunter & Furnham, 1984), the cultivation of fatalistic thoughts about cancer prevention from watching television news (Lee & Niederdeppe, 2011), the connection between narcissistic beliefs and preferences for specific genres of shows (Lull & Dickinson, 2016), the cultivation of fear and interpersonal trust across genres for Israeli youth (Cohen & Weimann, 2000), viewers’ perceptions of crime after watching a variety of genres (Grabe & Drew, 2007), and the impact of viewing different genres influences perceptions of sex within adolescents (Gottfried, Vaala, Bleakley, Hennessy, & Jordan, 2011) as well as other studies.

In introducing the genre variable into the study, there are difficulties in identifying particular characteristics about shows within a given genre. Consider situation comedies (sitcom), which has been one described as “one of the staples of mature broadcast television” (Hartley, 2001, pg. 65) but is exceedingly difficult to describe in a taxonomic categorization. An attempt at taxonomic categorization of sitcoms included details such as “fictional narrative,” “self-conscious performance,” “jokes and physical comedy,” and “the presence of a studio audience denoted by laughter on the soundtrack” (Bignell, 2013, pg. 133). Bignell (2013), only one sentence after outlining the characteristics seen in a sitcom, describes a show that would be considered a sitcom but does not fit those characteristics because it is shot on location rather than in front of a studio audience with a laugh track. Mills (2004) has gone so far as to state that that difference constitutes a new subgenre, dubbed by the author as “comedy vérité.”
Even within that subgenre, if one viewed it as such, one can outline characteristics that could separate programs that should be seemingly highly similar. One striking example of the differences between similar programs can be seen when comparing the British and American versions of the show *The Office* (Beeden & de Bruin, 2010). Beeden and de Bruin (2010) make the argument that national identity from the country of production introduces elements that influence how the audience will view the program, which in the case of *The Office* notably utilizes a different sense of humor and reference to cultural elements that are designed to be viewed as humorous and relevant by individuals from the same country where the respective shows were produced.

Given the previously discussed issues with genre definitions but impact of genre within media effects literature, it is beyond challenging to create specific hypotheses for how commercials may influence perceptions of program enjoyment across multiple (sub)genres. Therefore, two additional research questions have been added to the experiment:

RQ_{1a}: How will demographically-congruent commercials influence participants’ enjoyment of television shows across multiple genres?

RQ_{1b}: How will context-congruent commercials influence participants’ enjoyment of television shows across multiple genres?

Methodology

For Study 1, there were two experimental conditions. In one condition a “context congruent” group saw clips from three television shows targeted at senior citizens shown alongside advertisements targeted at that same age group. In the other condition a “demographically congruent” group saw the same clips from three television shows
targeted at senior citizens as the other group but with advertisements targeted at the participants’ age demographic. The three television shows selected can be classified as coming from three distinct and broadly defined genres: situation comedy (sitcom), drama, and news/talk. Episodes of The Golden Girls, Perry Mason, and The Florence Henderson Show were identified as sitcom, drama, and news/talk, respectively. The Golden Girls is a comedic show about four older, single/widowed women living together in Miami that aired on NBC from 1985-1992; Perry Mason is a dramatic show about a Californian defense attorney that aired on CBS from 1957-1966; The Florence Henderson Show is a news/talk show that features interviews with celebrities and informative clips various subjects that currently airs on RLTV (Retirement Living Television) and has since 2007.

To retain as much control as possible as well as continuity with former research in this domain (Saks, 2013), the same array of commercials used in prior research were also used in this study. Those commercials were selected from a pretest production-quality-evaluation study that included a convenience sample of 80 undergraduate college students aimed at maintaining that the audio/visual qualities of the two sets of commercials were not noticeably different (Saks, 2013). The “context congruent” commercials targeted at senior citizens included ads for adult diapers, reverse mortgages, retirement communities, and denture adhesive. The “demographically congruent” commercials targeted at college-aged individuals included ads for clothing, consumer electronics, and make-up.
The Implicit Association Test (IAT; Greenwald, McGhee, & Schwartz, 1998) (see Appendix A) used in this study attempts to see how quickly individuals can categorize lists of words that are either related to the stimuli or are “dummy nouns,” which are, in this case, colors (Wagner & Sundar, 2009). The participants categorized the stimuli-related terms and color names, along with positive and negative words, to examine the valenced strengths of association toward stimuli-related terms. The faster the individuals sort through the words associated with either the stimuli-related or the color name “dummy nouns,” when paired with either positive or negative words, across valence-paired trials, the more associated the stimuli-related terms are with positive and negative adjectives, subtracting the number of correctly-categorized stimuli-related terms, when paired with negative adjectives, from the number of stimuli-related terms, when paired with positive adjectives, across valence-paired trials. Participants categorized positive and negative adjectives with stimuli-related terms across two trials apiece. If a participant implicitly felt that the show was denotatively negative, one would more quickly associate words relating to the show with negative adjectives than with positive ones. Additionally, more words overall would be correctly categorized within 20 seconds during which one is tasked with identifying the words when the stimuli-related terms that signify the “attitude object” (Fazio, 1990) are categorized along with negative, as opposed to positive, words.

As mentioned previously, most studies have used minimal questionnaires to judge enjoyment amongst participants. Study 1 will derive its explicit measures from prior research (Palmgreen et al., 1991), as previous studies comparing implicit and explicit
attitudes have done (see, e.g., Andriasova & Wagner, 2004; Wagner, 2001; Wagner & Sundar, 2009; Saks, 2013). The explicit measure, as used in Study 1 and in the prior research from which it was adapted, includes eight 5-point semantic differential scales, including ranges from good to bad, acceptable to unacceptable, and wonderful to horrible in order to gauge participants’ overall attitudes toward the show (see Appendix B). Additional Likert scale statements were utilized for which the participants can either agree or disagree, asking participants to rate past viewing behaviors, liking of the show, liking of the commercials, and whether they would like to watch more of the show in the future. These types of measures have also been used in previous research (Raney & Bryant, 2002; Raney, 2005; Saks, 2013).

The participants engaged in the study in return for course credit, or had the option of doing an alternative assignment for the same credit in compliance with federal Institutional Review Board procedures. If they choose to participate in the experiment, once they arrived at the laboratory, they were given a consent form and were allowed to ask questions. If they consented to partake in the experiment, they were escorted into the lab where they were then randomly assigned to one of the two conditions. They were then connected to a measurement system consisting of a Biopac™ MP-35 unit connected to a pair of electrodes that were attached to their non-dominant hand, in order to measure arousal and do so with the least amount of time for participants to adapt to the electrodes or forget that they are strapped to their fingers (Sundar & Wagner, 2002). Then they were shown the stimulus to which they were randomly assigned. The order with which the
participants see the three stimuli for each condition was counterbalanced with an online random number generator so as to reduce the possibility of order effects.

Afterward, they first completed the implicit measure of association strength. Research (Converse, 1970; see Fazio & Towles-Schwen, 1999) has shown that filling out a questionnaire can actually lead participants to modify their responses and beliefs and/or create, on the spot, attitudes that they had not held prior to completing the questionnaire. The responses would be a function of completing an explicit measure, as opposed to demonstrating an implicit effect of the stimulus, which would invalidate the resulting implicit data should it be measured after filling out a self-report questionnaire. Therefore, the implicit attitude measure was conducted prior to the explicit measure. Once finished with the implicit measure, the participants were given the explicit measure. That process was then repeated twice more, with the participants seeing the next stimulus and completing the respective implicit and explicit measures for the videos in the other genres. At that point the experimental protocol was complete; the participants were then debriefed, thanked, and kindly escorted out of the lab.

Results

The first hypothesis focused on the explicit evaluations of the content by the participants. To test that hypothesis, after the participants viewed each stimuli they were asked to state their opinions of that specific clip across eight semantic differential scale items. Each item ranged from 1 to 5, with the lower numbers corresponding to more negative attitudes toward the program. After the participants evaluated the television clip the eight semantic differential items were then summed into an explicit index for each
clip. Each of the three indices was tested for reliability, with all being at acceptable levels (Golden Girls index: 8 items, $\alpha = .912$; Florence Henderson Show index: 8 items, $\alpha = .897$; Perry Mason index: 8 items, $\alpha = .853$). The range of the indexes was from 8 to 40, with 8 signifying the participant rating the clip as negatively as possible and with 40 signifying the participant rating the clip as positively as possible.

A total of 107 participants were recruited for the study. However, due to researcher error in terms of exhibiting the stimuli and multiple participants not showing up, the final number of people who participated in the study was 59. An a priori power analysis utilizing data from Saks (2013) showed that the minimum number of required participants per condition was 27. Due to random assignment, there were 28 participants in the “demographically congruent” group (which saw the shows for senior citizens paired with advertisements for college students) and there were 31 participants in the “context congruent” group (which saw the shows for senior citizens paired with advertisements for senior citizens).

The overall means (across both groups) for the indexes for the Golden Girls, Florence Henderson Show, and Perry Mason were 32.59, 27.49, and 27.85, respectively. For each video, the context-congruent group had a higher mean than the demographically-congruent group, but the differences were not significant when utilizing two-sided $t$-tests. In some cases, the means were almost identical, such as was the case for the Florence Henderson Show, where the context-congruent group had a mean of 27.58 (SD = 4.79), while the demographically-congruent group had a mean of 27.39 (SD = 5.43) [$t(57) = -.141, p = .888$]. Similarly, the context-congruent group for Perry Mason
(M = 28.10, SD = 4.73) was slightly greater than the mean for the demographically-congruent group (M = 27.57, SD = 4.32) \( t(57) = -0.444, p = .659 \). The mean indexes for the *Golden Girls* were the closest to producing significant results, with the context-congruent group’s mean of 33.39 and standard deviation of 5.40 being larger than the demographically-congruent group’s mean of 31.71 and standard deviation of 5.09 \( t(57) = -1.22, p = .226 \). Therefore, the first hypothesis was not supported.

Additional data that bolster the previously mentioned results come in the form of the participants’ responses to the following statements on a five-point Likert scale item, with responses ranging from strongly agree to strongly disagree: “Overall, I liked the television program I just watched” and “I would like to see more episodes of [title of the show] in the future.” Utilizing two-sided \( t \)-tests there were no significant differences amongst any of the comparisons for the two statements.

In an effort to support the second hypothesis, participants engaged in a word association task in the form of an Implicit Associations Task to gauge their internalized feelings about the shows. There are four critical stages of categorization: two in which the participants categorize the television show words and positive words together, and two in which the participants categorize the television show words and negative words together. To understand how the participants viewed the show, the correct number of associations made during the negative critical stages was subtracted from the correct number of associations made during the positive critical stages. Given that there were 8 television show words and 8 words for both positive and negative word types, the range of potential values for the implicit index was -32 to 32 across all four critical stages. An
A score of -32 would signify that the participant correctly associated all of the television show and negative words within the two negative critical stages without correctly associating any of the television and positive words during the two positive critical stages, and vice versa, for a score of 32.

Across the three television shows, when comparing the implicit index scores between the context-congruent and demographically-congruent groups, there were mixed results. For both *Perry Mason* and *The Florence Henderson Show* there was no significant difference between the implicit index scores when utilizing two-sided *t*-tests. *The Florence Henderson Show* mean scores for the demographically-congruent group (M = 1.04, SD = 8.02) was more positive than the mean scores of the context-congruent group (M = -0.42, SD = 8.61) [\(t(57) = .670, p = .51\)]. There was a similar difference for the *Perry Mason* mean scores, as the demographically-congruent group (M = -0.11, SD = 8.11) was more positive than the mean scores for the context-congruent group (M = -1.65, SD = 7.77) [\(t(57) = .744, p = .46\)]. However, there was a significant difference for the participants with regard to *The Golden Girls*. Those that saw the show paired with advertisements for college students (demographically-congruent) had a higher mean implicit index score (M = 5.75, SD = 7.37) than those that saw the show paired with advertisements for senior citizens (context-congruent) (M = 1.03, SD = 7.94) [\(t(57) = 2.357, p < .05\)]. As a result, the findings partially supported the second hypothesis.

**Discussion**

With neither of the study’s hypotheses fully supported, a variety of questions arise surrounding possible explanations for why that is the case. One potential explanation for
the difference between the hypotheses’ expected results and the study’s findings may come from the demographics of the content. As all three of the clips were targeted at individuals that are unlike the participants in the study they may have created separate effects. This was the main question at the heart of Study 1 – to examine if the effects from Saks (2013) remained intact outside of the audience’s demographic. As they did not, therein, it seems necessary to consider a variety of potential reasons for why those effects did not hold in Study 1.

One area to examine is that of the skin conductance levels (SCLs) of the participants while watching the content. As outlined, an essential element within Mandler’s (1991) theory is that there is an arousal response within the viewer that signifies that there is a discrepancy in the viewer’s understanding of what s/he should be seeing. When analyzing SCL the first step is to gauge a baseline for each participant’s tonic level (Potter & Bolls, 2012). In the case of this experiment, the main focus is on the commercial that is embedded within the program, as that is where the discrepancy should occur within the participants. To see if a physiological response occurred, various analyses were performed. For each video clip, there was one commercial break. The analyses included examining the skin conductance levels of the participants at the onset of the commercial break to those 5, 10, 15, and 20 seconds later, as well as the end of the commercial break. That suggests that the stimulus may not have been so sufficiently substantial as to cause a difference between the two groups, as well as it similarly suggests that the measures, as employed, were not adequately sensitive to demonstrate the hypothesized difference, or effect of the stimulus.
Additionally, a separate analysis compared the baseline measures for the participants at the beginning of each video to their levels at the end of each video. The results of all of these analyses can be seen in Table 1. It should be noted that the data for one participant in the context-congruent condition were lost due to researcher error. In all cases except two, (the 5 and 10 second marks after the commercial onset for *The Florence Henderson Show*), there was no significant difference in the participants’ SCLs. Therefore, those participants that viewed *The Florence Henderson Show*, paired with advertisements for senior citizens, experienced more physiological arousal within the first 10 seconds of the commercial break than those who saw the show with advertisements aimed at college students.

In reviewing that stimulus, it could be argued that the participants experienced a discrepancy, as the commercial in question advertised Depends undergarments for seniors. That commercial may have been exceedingly novel as compared to the others that the participants had viewed as part of the experiment. That commercial was included as part of Saks (2013), during which the participants did not experience a statistically significant physiological arousal. In contrast, when comparing the arousal levels between the two groups for *The Florence Henderson Show* across the entirety of the video, there was no significant difference. The discrepancy created may not have transferred the evaluation to the main content, in the form of either the implicit or explicit evaluations of the show.

As such, Mandler’s (1991) theory may not be ideal for examining these types of media. Despite the fact that Saks (2013) did find significant results for its explicit
evaluation hypothesis, the participants in that study did not have arousal responses to the content. Instead it may be necessary to utilize a different theoretical perspective to explain Saks’s (2013) results, which may ultimately lead to a better understanding of the results for Study 1 as well.

In comparing the results of the explicit evaluations of the shows to the results of Saks (2013), one should wonder why the previous study found a significant difference while the current study did not for its explicit evaluations hypothesis. There are a few differences in the two experiments that should be mentioned. First, Saks (2013) showed college-aged participants a single show clip that was targeted at their demographic paired with either commercials for that same college-aged demographic or commercials targeted at senior citizens. By showing three television show clips instead of the one that was exhibited in Saks (2013), the participants may have experienced fatigue by watching the clips and filling out both the implicit and explicit measures after each stimulus, as compared to going through the protocol only once. While the clips in Study 1 were randomly ordered so as to remove any potential side effects that would not eliminate the fatigue factor.

Table 1. Skin Conductance Levels by Group (A = Demographically-Congruent, B = Context-Congruent)

<table>
<thead>
<tr>
<th>Group</th>
<th>N</th>
<th>Mean</th>
<th>Std. Deviation</th>
<th>P-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Golden Girls Ad Onset To +5 seconds A</td>
<td>28</td>
<td>.001035430714286</td>
<td>.005587978588333</td>
<td>.439</td>
</tr>
<tr>
<td>B</td>
<td>30</td>
<td>.004272420333333</td>
<td>.021299845051364</td>
<td>.529</td>
</tr>
<tr>
<td>Golden Girls Ad Onset To +10 seconds A</td>
<td>28</td>
<td>.001689310357143</td>
<td>.007051884632303</td>
<td>.720</td>
</tr>
<tr>
<td>B</td>
<td>30</td>
<td>.004221613000000</td>
<td>.019998502351105</td>
<td>.720</td>
</tr>
<tr>
<td>Golden Girls Ad Onset To +15 seconds A</td>
<td>28</td>
<td>.003651157857143</td>
<td>.009354115372453</td>
<td>.720</td>
</tr>
<tr>
<td>B</td>
<td>30</td>
<td>.005137115000000</td>
<td>.019897047007048</td>
<td>.720</td>
</tr>
</tbody>
</table>
Table 1. Continued

<table>
<thead>
<tr>
<th>Table</th>
<th>Golden Girls</th>
<th>A</th>
<th>28</th>
<th>.003705639285714</th>
<th>.00891166761735</th>
<th>.528</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Ad Onset To</td>
<td>B</td>
<td>30</td>
<td>.001118963666667</td>
<td>.019728590194919</td>
<td></td>
</tr>
<tr>
<td></td>
<td>+20 seconds</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Table</td>
<td>Golden Girls</td>
<td>A</td>
<td>28</td>
<td>.007683793214286</td>
<td>.01619885623758</td>
<td>.887</td>
</tr>
<tr>
<td></td>
<td>Ad Onset To</td>
<td>B</td>
<td>30</td>
<td>.009002718000000</td>
<td>.046244614853725</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Ad End</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Table</td>
<td>Golden Girls</td>
<td>A</td>
<td>28</td>
<td>.109590792857143</td>
<td>.246349522661452</td>
<td>.152</td>
</tr>
<tr>
<td></td>
<td>Baseline</td>
<td>B</td>
<td>30</td>
<td>.236562091666667</td>
<td>.396243710731055</td>
<td></td>
</tr>
<tr>
<td>Table</td>
<td>Florence Henderson</td>
<td>A</td>
<td>28</td>
<td>-.001416804285714</td>
<td>.009088404892908</td>
<td>.041</td>
</tr>
<tr>
<td></td>
<td>Ad Onset To</td>
<td>B</td>
<td>30</td>
<td>.003102644333333</td>
<td>.007328735232498</td>
<td></td>
</tr>
<tr>
<td></td>
<td>+5 Seconds</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Table</td>
<td>Florence Henderson</td>
<td>A</td>
<td>28</td>
<td>.00218008214286</td>
<td>.009621021648429</td>
<td>.050</td>
</tr>
<tr>
<td></td>
<td>Ad Onset To</td>
<td>B</td>
<td>30</td>
<td>.005238861000000</td>
<td>.009429077423258</td>
<td></td>
</tr>
<tr>
<td></td>
<td>+10 seconds</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Table</td>
<td>Florence Henderson</td>
<td>A</td>
<td>28</td>
<td>.000545617857143</td>
<td>.01513645005461</td>
<td>.158</td>
</tr>
<tr>
<td></td>
<td>Ad Onset To</td>
<td>B</td>
<td>30</td>
<td>.004628462666667</td>
<td>.00848611075151</td>
<td></td>
</tr>
<tr>
<td></td>
<td>+15 seconds</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Table</td>
<td>Florence Henderson</td>
<td>A</td>
<td>28</td>
<td>.002724813214286</td>
<td>.01661342986053</td>
<td>.391</td>
</tr>
<tr>
<td></td>
<td>Ad Onset To</td>
<td>B</td>
<td>30</td>
<td>.005900087666667</td>
<td>.010972151143652</td>
<td></td>
</tr>
<tr>
<td></td>
<td>+20 seconds</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Table</td>
<td>Florence Henderson</td>
<td>A</td>
<td>28</td>
<td>.008283377857143</td>
<td>.04031876856472</td>
<td>.813</td>
</tr>
<tr>
<td></td>
<td>Ad Onset To</td>
<td>B</td>
<td>30</td>
<td>.006454952500000</td>
<td>.011388262796815</td>
<td></td>
</tr>
<tr>
<td></td>
<td>+10 seconds</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Table</td>
<td>Perry Mason</td>
<td>A</td>
<td>28</td>
<td>-.000490504285714</td>
<td>.005975774435520</td>
<td>.654</td>
</tr>
<tr>
<td></td>
<td>Ad Onset To</td>
<td>B</td>
<td>30</td>
<td>.002034130000000</td>
<td>.005747578439915</td>
<td></td>
</tr>
<tr>
<td></td>
<td>+5 seconds</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Table</td>
<td>Perry Mason</td>
<td>A</td>
<td>28</td>
<td>.002070792500000</td>
<td>.007947363244831</td>
<td>.369</td>
</tr>
<tr>
<td></td>
<td>Ad Onset To</td>
<td>B</td>
<td>30</td>
<td>.000356012333333</td>
<td>.00644893959798</td>
<td></td>
</tr>
<tr>
<td></td>
<td>+10 seconds</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Table</td>
<td>Perry Mason</td>
<td>A</td>
<td>28</td>
<td>.001743765714286</td>
<td>.00787638772394</td>
<td>.387</td>
</tr>
<tr>
<td></td>
<td>Ad Onset To</td>
<td>B</td>
<td>30</td>
<td>-.000101743000000</td>
<td>.008231099487430</td>
<td></td>
</tr>
<tr>
<td></td>
<td>+15 seconds</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Table</td>
<td>Perry Mason</td>
<td>A</td>
<td>28</td>
<td>.002234351428571</td>
<td>.011239649508361</td>
<td>.699</td>
</tr>
<tr>
<td></td>
<td>Ad Onset To</td>
<td>B</td>
<td>30</td>
<td>.003356916666667</td>
<td>.010740392698065</td>
<td></td>
</tr>
<tr>
<td></td>
<td>+20 seconds</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Table</td>
<td>Perry Mason</td>
<td>A</td>
<td>28</td>
<td>.006757380714286</td>
<td>.016861097435404</td>
<td>.221</td>
</tr>
<tr>
<td></td>
<td>Ad Onset To</td>
<td>B</td>
<td>30</td>
<td>.002396270000000</td>
<td>.009531072065840</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Ad End</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Table</td>
<td>Perry Mason</td>
<td>A</td>
<td>28</td>
<td>.012534080357143</td>
<td>.077827330762893</td>
<td>.730</td>
</tr>
<tr>
<td></td>
<td>Base To End</td>
<td>B</td>
<td>30</td>
<td>.020446755000000</td>
<td>.094373855887826</td>
<td></td>
</tr>
</tbody>
</table>

Another major methodological difference between Saks (2013) and Study 1 is that Saks (2013) utilized two commercial breaks instead of only one per clip in an attempt to balance the need to show all of the materials to the participants while also keeping the
clips short enough to avoid fatigue. In hindsight it would have been better to keep multiple commercial breaks to maintain continuity with Saks (2013), as well as to potentially strengthen the stimulus so sufficiently as to yield the hypothesized effect. As such, future replications should include two commercial breaks to increase the strength and salience of the manipulation as well as adhere to closer similarity with Saks (2013). Additionally, given the necessary resources and time, the experimental protocol could be broken down to show the participants only one clip, instead of three, so as to limit the potential effects of fatigue.

From a theoretical perspective, not having found more than one significant result in Study 1 creates additional questions as to what may have caused the effect seen in Saks (2013). If the results from Saks (2013) were in fact an accurate representation of a phenomenon involving the influence of television commercials upon viewers’ perceptions of television shows, it may be necessary for that content that is enjoyable for a given audience. Considering that both experiments used age-based incongruity, potential future studies could test other forms of incongruity to see if correctly-targeted (or congruently-targeted) content is essential for the commercials to have an influence.

As a function of Study 1’s results, as well as those of prior work by Saks (2013), a follow-up study should allow for the resolution of the discrepancies between theory and the results of those studies’ theory-based hypothesis findings. Toward that end, Study 2 attempts to create an environment where the participants’ feel that the content is directly tailored to their preferences, as opposed to having been inferred from the preferences of similar individuals within their age group. In doing so, the potential influence of outside
information on the participants’ perceptions of the content may be more noticeable, resulting in statistically-significant results that (more) fully support the forwarded, theory-based hypotheses.
STUDY 2: EFFECTS OF CUSTOMIZED RATINGS ON USER EVALUATIONS OF TELEVISION SHOWS

Literature Review

Although several studies have investigated the use of personalized content recommendation/rating services (Liang, Lai, & Ku, 2006; Wessel & Thies, 2015; Chau, Ho, Ho, & Yao, 2013), not much, if any, research has examined the way customized ratings influence users' enjoyment of media. Liang, Lai, and Ku’s (2006) research, which more specifically examined user satisfaction as it relates to customized online news, as compared to, for instance, gratification, hence likely measured what could be termed “usefulness” more than what could be argued to be “enjoyment,” per se. Wessel and Thies (2015) also focus on a personalized news service, but they determine intent to buy such a service, also a measure of usefulness. The majority of the studies regarding customized ratings are related to the computer science and engineering aspects of creating and exhibiting the information (Claypool, Le, Wased, & Brown, 2001; Pu, Chen, & Hu, 2011; Dooms, De Pessemier, & Martens, 2011). It is notable to point out that in many of the studies (Häubl & Murray, 2003; Gretzel & Fesenmaier, 2005; Cremonesi, Garzotto, & Turrin, 2012; for an extensive literature review of recommendation systems within the television domain, see Véras, Prota, Bispo, Prudêncio, & Ferraz, 2015), the authors utilized their own recommendation systems that yielded different outcome stimuli based on the individual responses of the participants. In doing so, the studies are assuming that the very act of giving the participants recommendations does not alter their perceptions of the target content. Therefore, in actuality, those studies are ultimately
evaluating their specific recommendation algorithms instead of the total potential for any recommendation to alter the participants' evaluations of a target stimulus. Knijnenburg and associates (2012) does make mention of the potential for recommendation systems to result in behavioral outcomes, and as such leaves open various questions on what those outcomes are, how they occur, when they occur, under what circumstances they occur, and what are the underlying mechanisms related to them.

Other research has analyzed how similarly focused content, namely movie reviews, may alter viewers’ evaluations (Chakravarty, Liu, & Mazumdar, 2010). The authors found reviews and electronic word-of-mouth (eWOM) descriptions of not-yet-released movies could sufficiently manipulate the participants’ interest in the target movie. That study did not actually show the participants the movie, however, and utilized within-group manipulations. Additional research (d'Astous & Touil, 1999; d'Astous & Colbert, 2002) does offer evidence to buttress those claims.

One factor to consider when evaluating how viewers will rate content is their prior expectations generated toward the content (Cardozo, 1965; Anderson, 1973). Expectancy-Disconfirmation theory (alternatively referred to as “Expectancy-Confirmation” theory), as constructed through multiple studies (Cardozo, 1965; Anderson, 1973; Oliver, 1980), states that consumers’ attitudes about a product will be shaped by their previous understanding of the qualities of that product prior to experiencing it. Once they experience the product, or media content the users will compare their actual experience to that of their expectations. Therefore, if the product is
of equal or higher quality than that originally imagined, a consumer will be pleased, whereas if the product is of lower quality than anticipated, a consumer will be displeased.

In the case of the online media distributors who include customized ratings and suggestions, it is proposed that the ratings that the individuals receive prior to watching the content will produce the same types of expectations. Some participants may also have certain expectations about categories of content, such as genres, based on prior knowledge and experiences (Sujan, 1985); however in Study 2, random assignment should control for those previous notions about specific types of content.

Research utilizing Expectancy-Disconfirmation theory covers a variety of topics, although the majority of it is related to consumers in multiple different contexts. The theory is heavily utilized in marketing research to examine individuals’ purchase and repurchase intentions in consideration of their initial expectations as well as the individuals’ experiences of the item/service after the fact (Hossain & Quaddus, 2012).

Such studies have analyzed topics including service quality and wait time at restaurants (Davis & Heineke, 1998; Oh, 1999), satisfaction among tourists (Pizam & Milman, 1993; Zehrer, Crotts, & Magnini, 2011; Hui, Wan, & Ho, 2007), repurchase of an automobile (Oliver, 1993), and purchase of photographic and video recording products (Spreng, MacKenzie, & Olshavsky, 1996; Dabholkar, Shepherd, & Thorpe, 2000). The theory has also been utilized for research on media. Tsuji, Bennett, and Zhang (2007) examined user satisfaction with an action sports event, Jacobs (1995) explored the factors in understanding subscribers’ satisfaction with their cable television subscription, and
Ladhari (2007) utilized to theory to understand the outcomes for moviegoers’ experiences.

Other avenues of research within the framework have evaluated issues unrelated to consuming products and services. Such foci include changes in marital relationship (Hackel & Ruble, 1992), experiencing an encounter with police (Reisig & Stroshine Chandek, 2001), and understanding a citizen’s view of their government (Van Ryzin, 2006). Although satisfaction was widely included as a key dependent variable(s) in many of the studies employing Expectancy-Disconfirmation theory, enjoyment (the current focus) was not utilized often. However, multiple studies have identified positive affective response as a precursor for individuals feeling satisfied while also having negative affective response leading to a feeling of dissatisfaction (Oliver, 1993; Mano & Oliver, 1993; Westbrook, 1987; Westbrook & Oliver, 1993).

An additional element within that theoretical perspective that requires discussion is priming (Meyer & Schvaneveldt, 1971; Iyengar, Peters, & Kinder, 1982). The experience of seeing the rating prior to viewing the material should prime participants to similarly or conversely evaluate the content, as a function of the extremity of the rating, from low to high, respectively. Repetitive priming has the potential to increase automatic retrieval and processing speed (Logan, 1990), and the ease with which individuals can process information related to an attitude-evoking stimulus has the potential to influence its visual attractiveness as opposed to other stimuli (Roskos-Ewoldsen & Fazio, 1992), which help form the mechanisms that underlie expectancy (dis)confirmation.
Research (Oliver, 1977) has shown that the priming of a baseline measure within an experimental context, such as the (experimenter-manipulated) customized ratings given to the participants in Study 2, can shift post-stimulus evaluations with regard to that baseline. If the ratings are sufficient to create a preconceived notion, the results should show a difference between those who were primed to show positive levels of enjoyment toward a target stimulus and those who were primed to show negative levels of enjoyment toward a target stimulus.

As a function of participants being primed with different content ratings, in order to cause (dis)confirmation of the expectations theorized to be formed in processing different ratings, two sets of hypotheses for Study 2 can be forwarded as a result of measuring participants’ levels of target stimulus enjoyment in two different ways:

H\textsubscript{1a}: The participants in the high-rating group will rate the stimulus to be more enjoyable on the seven-point rating scale than the low-rating group.

H\textsubscript{1b}: The participants in the low-rating group will rate the stimulus to be less enjoyable on the seven-point rating scale than the high-rating group.

H\textsubscript{2a}: The participants in the high-rating group will rate the stimulus to be more enjoyable across the totality of eight semantic differential items than the low-rating group.

H\textsubscript{2b}: The participants in the low-rating group will rate the stimulus to be less enjoyable across the totality of eight semantic differential items than the high-rating group.

RQ\textsubscript{1a}: How will participants’ enjoyment of content change when primed with high-ratings?

RQ\textsubscript{1b}: How will participants’ enjoyment of content change when primed with low-ratings?
Methodology

Experimental Design

Study 2 is a three-condition design utilizing expectancy disconfirmation as the main theoretical basis for testing the influence of customized, “Netflix-like” ratings on users’ evaluations of content. There will be three treatments: a high rating, a low rating, and a control. The control group is used to examine whether there is a difference between each the high and low rating groups and not having seen any ratings prior to viewing the target stimulus—or rather, a baseline score—in order to examine whether processing ratings can change in participants’ level of enjoyment, as opposed to simply being able to test whether there is a difference between being primed with high versus low ratings.

At the beginning of the study sessions, participants will be told that they are to rate three target stimuli. They will also be told that they will first need to rate other media content to provide a baseline reference for their eventual rating of the three targets. After they rate the initial media content and before viewing the target stimulus, the system will provide a supposedly customized rating of how much the experimental groups will enjoy the target stimuli. The “customized” output ratings, however, will be independent of the individual user’s inputted ratings and will be consistent for all participants in each treatment. Those in the high rating group will be told that the algorithm has found that they will evaluate the following programs highly on a seven-point scale, while those in the low rating group will be told that they will evaluate the content poorly on the same seven-point scale. The control groups will not be given a customized rating. Therefore, given Expectancy-Disconfirmation theory, the ratings
should act as the priming source of information through which the participants create the preconceptions about the stimuli.

**Sampling**

The participants for this experiment include a convenience sample from the E.W. Scripps Participant Pool. The demographics for these participants are college students in an introductory strategic communication course. They are typically between 18 and 23 years old. They are almost entirely from the United States, and if not, it is assumed that they are studying in the U.S. at least in part because they are interested in American culture. All students receive partial course credit for participating in the study or for fulfilling another comparable assignment in compliance with U.S. federal Institutional Review Board procedures. A total of 117 individuals participated in the study. The participants were randomly assigned to the treatments when they arrived at the laboratory with 39 (33.3%) in the high rating group, 40 (34.2%) in the low rating group, and 38 (32.5%) in the control group.

**Stimulus Design**

The stimulus materials for Study 2 were selected from multiple genres of television programs and to appeal to the demographics of the participant while also being episodes of shows that have ideally not been seen by the participants prior to the experiment. The three television shows that were selected were from situation comedy, drama, and news genres, including *Better Off Ted*, *Studio 60 on the Sunset Strip*, and *TechKnow*, respectively. *Better Off Ted* was a workplace comedy that originally aired on ABC for two seasons from 2009-2010 and is currently available on Netflix; *Studio 60 on
the Sunset Strip was a dramatic series focusing on the staff of a late-night comedy show created by Aaron Sorkin and aired for one season on NBC from 2006-2007 which is available for purchase on an individual episode basis from various online outlets; TechKnow is a news/information show about technology and science that aired from 2013-2016, when Al Jazeera English ceased broadcasting entirely. (Full episodes are available through the former channel’s website.) Shows with older original dates and short spans during which the shows originally aired were chosen to help ensure that participants had not seen them, and while both of those factors could be argued to mean that the shows were not originally enjoyed by the target audience, if that is the case, then it would lower the likelihood of finding statistically significant results for the low rating hypotheses and research questions about changes in levels of enjoyment, thereby making findings regarding such hypotheses and dependent variable level changes about any (media-related) stimuli stronger (see, e.g., Wagner, 1998; Sundar, 1999). The materials were downloaded in the highest quality possible digital video file.

The experiment took place on four computers utilizing MediaLab software. Along with the target materials, the participants also rated other media content so that the cover story of personalized ratings is seemingly valid. A total of 65 popular television shows were cataloged, with relevant promotional material (such as posters, cast photos, or title screens) gathered, so that the participants could judge how they liked them. To ensure that many of the participants will have seen a substantial proportion of the material available for judgment, the relatively large number of promotion materials were cataloged. The participants then judged the shows represented by the materials on the
same seven-point scale on which they later gave the customized rating if they were not in the control condition. In addition to the seven-point scale, there was a separate option for the participants to state that they have not seen the media for which the promotional materials were related.

**Measurements / Instruments**

There is one independent variable for Study 2: the “customized ratings” given to the participants. The dependent variables will include the overall rating that the participant gives to the target stimulus on a seven-point scale and their answers to eight semantic differential items used in previous studies (Saks, 2013; Palmgreen et al., 1991; Wagner, 2001; Wagner & Sundar, 2009). The semantic differential items, available in Appendix C, will then be summed to create an index. The answers from the individuals in the treatment groups will be compared with those in the control group to see what the baseline, unprimed responses are to the material.

The study also includes manipulation checks and control variables (see Appendix D). For manipulation checks, after viewing the target stimulus, the participants will be asked if they have ever heard of or seen the show prior to the experiment. Additionally, the initial inputted rankings for each media content will be kept and analyzed to understand if there are any patterns that led to specific results, such as a large number of “have not seen” responses. Control variables include understanding the demographic characteristics of the participants, including gender and nationality. Some materials may also skew toward individuals that identify with specific genders. Finally, it is important
to know from where the participants hail so as to know if there are any cultural factors that may influence their preference for particular media.

Overall, it should be kept in mind, regarding the above description of the sample, that random assignment of participants to conditions control for any and all individual differences among participants across groups and that any discrepancies between the hypothesized effects and realized effects, as a function of differences between the idealized groups of participants and those who took part in the study, based on the demographic composition of the actual convenience sample would only serve to strengthen significant differences, or effects, found by the study (see Wagner, 1998; Sundar 1999).

Procedure

Once the participants arrived at the laboratory they were given a consent form and the opportunity to ask questions. If they consented to the experiment they were escorted into the lab where they pulled an ID number out of an opaque bag/envelope. The ID assigned each respective individual into one of the three potential groups. S/he was then seated at a private computer terminal that would administer the treatment. Up to four individuals participated in the study at one time, but the lab was set up in such a way that each person could only see one’s own monitor during the protocol reading. The participants were then told that they were going to test a new customized rating system, similar to Netflix. In order to create the customized rating (according to the backstory of the experiment), they first gave their ratings on other material. Titles and promotional materials for the cataloged media examples were displayed, one at a time, and the
participants rated each one on a seven-point scale or stated that they have not seen it. After they had completed those ratings, those not in the control condition were shown a rating for how the algorithm found they would like the following program. Participants in the control condition were told that they would see their customized rating after watching the stimuli. The participants then watched the target stimuli and were instructed that they would be asked their opinions on the content over a few questions and that they should answer them as quickly and accurately as possible. The questions and statements in Appendices C and D were administered to the participants. At this point, the participants were debriefed, thanked for their participation, and told that they can leave the laboratory or ask questions.

Results

For Study 2, a total of 117 participants were recruited. Due to random assignment, 39, 40, and 38 participants were assigned to the positive, negative, and control conditions, respectively. The first set of hypotheses for Study 2 deals with the participants’ evaluations of the television clips on the same seven-point scale used to prime their expectations. The overall means for the three shows were neutral to slightly positive with Better Off Ted having a mean of 3.79 and standard deviation of 1.36, Studio 60 on the Sunset Strip having a mean of 4.86 and standard deviation of 1.74, and TechKnow having a mean of 4.26 and a standard deviation of 1.67.

Using a one-way analysis of variance (ANOVA) for each of the three television clips, all three clips were significant: Studio 60 \( F(2, 114) = 3.52, p < .05 \), TechKnow \( F(2, 114) = 3.16, p < .05 \) and Better Off Ted \( F(2, 114) = 6.17, p < .01 \). When using
Tukey HSD post-hoc tests, there is a significant difference for all three shows when comparing the positive condition to the negative condition. The difference between the positive and negative conditions for Better Off Ted was significant at $p < .01$, while the other two shows were significant at the $p < .05$ level. For the full list of means and standard deviations for all three shows, see Table 2. Given these data, $H_{1a}$ and $H_{1b}$ are supported.

Table 2. Mean and Standard Deviation on Seven-Point Scale by Condition for Each Show

<table>
<thead>
<tr>
<th>Show</th>
<th>Condition</th>
<th>N</th>
<th>Mean</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Better Off Ted</td>
<td>Positive</td>
<td>39</td>
<td>4.31**</td>
<td>1.26</td>
</tr>
<tr>
<td></td>
<td>Negative</td>
<td>40</td>
<td>3.28**</td>
<td>1.30</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>38</td>
<td>3.79</td>
<td>1.36</td>
</tr>
<tr>
<td>Studio 60 on the Sunset Strip</td>
<td>Positive</td>
<td>39</td>
<td>5.33*</td>
<td>1.58</td>
</tr>
<tr>
<td></td>
<td>Negative</td>
<td>40</td>
<td>4.33*</td>
<td>1.91</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>38</td>
<td>4.95</td>
<td>1.74</td>
</tr>
<tr>
<td>TechKnow</td>
<td>Positive</td>
<td>39</td>
<td>4.64*</td>
<td>1.75</td>
</tr>
<tr>
<td></td>
<td>Negative</td>
<td>40</td>
<td>3.75*</td>
<td>1.51</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>38</td>
<td>4.42</td>
<td>1.65</td>
</tr>
</tbody>
</table>

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

The second set of hypotheses is similar to the first, however it compares the participants’ evaluations across eight semantic differential scale items. The eight items each ranged from 1 to 5, with 1 being the negative side of the scale and 5 being the positive side. Those items were then summed to create an index with potential values ranging from 8 to 40. Each of the scales for the three shows were at acceptable levels of
reliability (*Better Off Ted* index: 8 items, $\alpha = .959$; *Studio 60* index: 8 items, $\alpha = .979$; *TechKnow* index: 8 items, $\alpha = .965$). The overall means and standard deviations for the shows (regardless of condition) were as follows: *Better Off Ted* ($M = 24.85, SD = 7.05$), *Studio 60* ($M = 29.09, SD = 8.49$), and *TechKnow* ($M = 28.50, SD = 8.31$).

When utilizing one-way ANOVAs to test the three shows for differences across all conditions, each one was significant: *Studio 60* [$F(2, 114) = 3.68, p < .05$], *TechKnow* [$F(2, 114) = 3.82, p < .05$] and *Better Off Ted* [$F(2, 114) = 4.66, p < .05$]. Utilizing Tukey’s HSD post-hoc tests, there were four significant differences among all three videos. Those differences were between *Better Off Ted*’s positive ($M = 27.33, SD = 6.94$) and negative ($M = 22.65, SD = 6.10$) conditions ($p < .01$), *Studio 60*’s positive ($M = 31.26, SD = 8.52$) and negative ($M = 26.32, SD = 9.08$) conditions ($p < .05$), *TechKnow*’s positive ($M = 29.97, SD = 8.26$) and negative ($M = 25.62, SD = 8.32$) conditions ($p < .05$), and *TechKnow*’s negative ($M = 25.62, SD = 8.32$) and control (30.02, SD = 7.75) conditions ($p < .05$). All of the means and standard deviations across all three videos for all three conditions can be seen in Table 3. Given these data, H$_{2a}$ and H$_{2b}$ are both supported.

Along with the seven-point scale and indexes for each of the shows, the participants were asked to respond to three additional five-point Likert scale items (see Appendix D), with 1 corresponding to “strongly disagree” and 5 corresponding to “strongly agree.” Those items asked the participants to judge their overall liking of the show, their past experience with the show, and if they would like to see more of the show in the future. For the first and third statements, the participants stated evaluations were
similar to the previously two discussed measures, except few of the comparisons were
significant. The only statement which had significantly different results when analyzed
with a one-way ANOVA was for Better Off Ted, for the statement “Overall, I liked the
television program I just watched” \( [F(2, 114) = 3.999, p < .05] \). According to a Tukey’s
HSD post-hoc test, for that statement the positive condition (M = 3.31, SD = 0.98) was
significantly greater than both the negative condition (M = 2.75, SD = 1.01) and the
control condition (M = 2.76, SD = 1.00).

Table 3. Mean and Standard Deviation of Indexes by Condition for Each Show

<table>
<thead>
<tr>
<th>Show</th>
<th>Condition</th>
<th>N</th>
<th>Mean</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Better Off Ted</td>
<td>Positive</td>
<td>39</td>
<td>27.33</td>
<td>6.94</td>
</tr>
<tr>
<td></td>
<td>Negative</td>
<td>40</td>
<td>22.65</td>
<td>6.10</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>38</td>
<td>24.60</td>
<td>7.44</td>
</tr>
<tr>
<td>Studio 60 on the Sunset Strip</td>
<td>Positive</td>
<td>39</td>
<td>31.26</td>
<td>8.52</td>
</tr>
<tr>
<td></td>
<td>Negative</td>
<td>40</td>
<td>26.32</td>
<td>9.08</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>38</td>
<td>29.79</td>
<td>7.11</td>
</tr>
<tr>
<td>TechKnow</td>
<td>Positive</td>
<td>39</td>
<td>29.97</td>
<td>8.26</td>
</tr>
<tr>
<td></td>
<td>Negative</td>
<td>40</td>
<td>25.62</td>
<td>8.32</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>38</td>
<td>30.02</td>
<td>7.75</td>
</tr>
</tbody>
</table>

\( a \ p < .01, \ b \ p < .05, \ c \ p < .05, \ d \ p < .05 \)

Despite the fact that only one ANOVA was significant, the responses to both
statements for all three shows had a similar pattern. The positive condition participants’
responses were higher than the control condition participants’ responses, which were in
turn higher than the negative condition participants’ responses, except in the case of
TechKnow, where the control response was the highest, followed by the positive condition, and then the negative condition.

In the case of their past experiences with the shows, both Better Off Ted \(F(2, 114) = .140, p = .840\) and Studio 60 \(F(2, 114) = 1.479, p = .232\) had no significant differences across all of the conditions with a one-way ANOVA or when compared individually on a condition-by-condition basis utilizing a Tukey’s HSD post-hoc test. However, TechKnow \(F(2, 114) = 4.447, p < .02\) did have a significant difference in the number of individuals that stated they watched the often when analyzed with a one-way ANOVA. That significant difference occurred, according to a Tukey’s HSD post-hoc test, between the positive (M = 2.26, SD = 1.27) and negative (M = 1.55, SD = 0.85) conditions \(p < .05\). A detailed breakdown of the participants’ responses to the statement “I watch (program name) often” can be seen in Table 4. However, there are a variety of reasons why the participants’ responses to this question may not be valid. Those reasons will be highlighted in the Discussion section for Study 2.

<table>
<thead>
<tr>
<th></th>
<th>Better Off Ted</th>
<th>Studio 60</th>
<th>TechKnow</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pos</td>
<td>Neg</td>
<td>Cont</td>
</tr>
<tr>
<td>Strongly Disagree</td>
<td>19</td>
<td>25</td>
<td>21</td>
</tr>
<tr>
<td>Disagree</td>
<td>11</td>
<td>7</td>
<td>12</td>
</tr>
<tr>
<td>Neutral</td>
<td>9</td>
<td>6</td>
<td>3</td>
</tr>
<tr>
<td>Agree</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Strongly Agree</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Total</td>
<td>39</td>
<td>40</td>
<td>38</td>
</tr>
</tbody>
</table>
Discussion

Given the support for the hypotheses, Study 2 makes a number of contributions to knowledge about understanding how personalization of content may influence individuals’ evaluations. Personalization and targeted recommendations are a common part of the modern media environment; however the effects on the users have not been studied in a similarly controlled way. It should give researchers and the media industry pause to know that these recommendations may impact how the content is viewed.

Considering the research on recommendation systems that has already been completed (Häubl & Murray, 2003; Gretzel & Fesenmaier, 2005; Cremonesi, Garzotto, & Turrin, 2012; Véras, Prota, Bispo, Prudêncio, & Ferraz, 2015), researchers should want to know, from both a theoretical and methodological standpoint, whether simply making a suggestion can influence the results of their studies. Additionally, it could be hypothetically possible that the content providers utilize the feedback from users in determining what types of content they will ultimately create or support. Should that be the case, those providers would want to understand what impact the presence of a rating system has on the user. That knowledge could be especially useful for the content exhibitors that are also content creators, which includes Netflix, Amazon Prime, and Hulu. It could be hypothetically possible, if the effects shown in Study 2 hold true outside of the laboratory environment, that those exhibitors could manipulate ratings to increase viewers’ liking of the show. Considering that the exhibitors have a vested interest in maintaining quality content so that their subscribers will see value in the subscription fee, simply manipulating the rating to be higher could prime the viewers to
like the content. Viewers may therein decide that the service is more worthy of their disposable income.

Users of such services could use the findings outlined in Study 2 to better understand how they may be manipulated without their knowledge. Further research should examine if the influence can be limited if participants are informed of the potential effect prior to viewing the stimuli. As more websites and services utilize algorithmic-based content ratings, recommendations, and disseminations, the potential influence on the users grows. Such influence can be seen in how individuals get and receive news from the algorithmic-based website, Facebook, and youth’s shifting preferences for differing types of news (Marchi, 2012).

When looking at both sets of hypotheses together, the results of Study 2 offer support that the act of giving a recommendation may influence the views of those that receive that recommendation. Not only were the evaluations significantly greater for those that received higher priming scores on the seven-point scale, those primed expectations also transferred to the participants evaluating the content differently across the index of the eight semantic differential scale items. For all three of the videos on the original seven-point scale (Figure 1) and two of the three videos on the eight-item indexed scale (Figure 2), the mean responses for the positive condition groups were higher than that of the control, which were in turn higher than the mean responses of the negative condition group. For TechKnow, on the eight-item indexed scale, the participants in the control condition group had a slightly higher mean response \( M = 30.03 \) than the positive condition group \( M = 29.97 \). Despite the seeming trend of the
positive condition being higher than the control, it is not possible to state that that is due to the manipulation, as that was not statistically significant in any of the analyses. However, it is possible to state that the manipulation did create a difference between the positive and negative conditions, as those groups were statistically significantly different from one another in all six of the analyses.

Figure 1. Mean Response to the Videos by Condition on Seven-Point Scale
Another consideration when looking at the results with regard to the effect sizes comes from the understanding of media enjoyment as utilized in this study. There are numerous different ways in which media enjoyment has been defined and operationalized, leading to an incomplete and scattered view of the phenomena. The understanding of enjoyment utilized for the current purposes, which came from Nabi and Krcmar (2004), broke the concept down into three parts: affective, cognitive, and behavioral. Research attempted to measure each element within that concept by considering either the ratings given by the participants (cognitive) or their responses to the statements about liking the show (affective). It also considered if viewers wanted to see that show again in the future (behavioral). There was significance for the cognitive elements for all of the shows across the conditions when comparing the positive to the
negative conditions as well as the control condition when compared to the negative condition for *TechKnow*. However there were not many significant effects seen when analyzing the affective or behavioral dimensions. Only *Better Off Ted* had a significant difference in terms of the affective component, while there was no difference among the conditions for any of the shows for the behavioral dimension. Given that those dimensions were only being measured by a single question each, they may not have fully captured the scope of the influence on the participants. In all of the cases, the mean participant responses for the two questions mirrored their responses to the mean responses for the cognitive dimension, with the positive conditions being greater than the control, which was greater than the negative. This was not found in the case of *TechKnow*, where the control condition mean response was the greatest. The mean participant responses for the affective and behavioral statements by condition can be seen in Figure 3.
It is important to consider those data when thinking about the overall concept of enjoyment for the participants. It is possible to conclude that one dimension of the concept was manipulated by the ratings, however not all elements were influenced equally. Other studies that employ a similar three-pronged conceptualization of enjoyment have utilized more elements to measure each dimension (Krakowiak & Oliver, 2012; Raney & Bryant, 2002), however those studies were focusing on longer and more in-depth narratives. Given the short length of the clips in the present study, the participants would likely not have had the time necessary to form judgments related to elements that other studies highlighted, such as the moral inclinations of characters. As such, the cognitive elements for the current study are underdeveloped. Future studies should take more time with a singular media text or series of texts from the same show.
and examine the more nuanced details that are not captured in a 10-minute clip. Those cognitions would also be different based on the expectations of genres, noting that one may enjoy a news program and a comedy program equally, but in different ways given Uses and Gratifications (a separate sub-paradigm within media effects research that examines how and why individuals use specific media) (see Lazarsfeld, 1941) for those specific programs and genres. With that in mind, it is notable that the affective dimension for *Better Off Ted* (comedy) was significant \[ F(2, 114) = 3.999, p < .05 \]. Therefore there may have been an element within the way the participants enjoyed the comedy genre that led to the affective response for it, as compared to the two other genres.

In considering enjoyment, one should recognize that some aspects of enjoyment may fluctuate over time and are difficult to measure. Nabi and Krcmar note, “Indeed, as the receiver reflects on the message long after exposure, expressions of enjoyment, like attitudes, can change even still (again, likely a function of changes in the weighting of the underlying components)” (2004, p. 295). They continue, “Enjoyment has an experiential component that does not, at first, appear to be well captured by the notion of attitude. That is, though one may have a favorable or unfavorable attitude toward an object, there is no implication that this is a summary of a broader experience” (Nabi & Krcmar, 2004, pp. 295-296). It may require time to more fully understand the media enjoyment experience for the participants. Examining results over time with consideration for cultivation theory may lead to better understanding of the overall concept of media enjoyment.
Given that the three shows came from three different genres, it is notable that the
participants still were influenced by the ratings in most of the cases. Much of the
literature on genre-related media effects is focused on cultivation theory, which considers
the slow and steady growth of effects over time, such as believing the world is more
violent than it actually is after consistently watching television news (Gerbner & Gross,
1976). However, in the case of Study 2, the short length of the clips with the relatively
minor priming by the independent variable would likely not be considered consistent
enough to warrant examination within the parameters of cultivation research. As such,
the genre variable does not appear to influence the effectiveness of the treatment. It
should be noted that for both hypotheses the analyses did not have large effect sizes. The
largest effect size was found on the one-way ANOVA for Better Off Ted (comedy) on the
original seven-point scale ($\eta^2 = .098$), with the smallest found on the one-way ANOVA
for TechKnow (news/information) on the seven-point scale ($\eta^2 = .052$).

Additional variables could be used to better understand the full scope of variance
occurring within the effects. One variable not included in this study was perceived trust
in the recommendation agent. Studies, mostly within computer science and information
systems, have included trust in the recommendation agent as both an independent and
dependent variable (Gefen, Karahanna, & Straub, 2003; Urban & Hauser, 2004;
Jarvenpaa, Tractinsky, & Saarinen, 1999; Cremonesi, Garzotto, & Turrin, 2012). By
including that variable in future studies one could better understand if the participants
truly believed that the recommendations were targeted at them.
One potential explanation for the differing findings and effect sizes for *TechKnow* may come from the types of students that were participating in the study. Many of the students who participated in the study may have been members of the journalism department, despite the fact that they were self-selected from a strategic communication course. Given departmental procedures, all students who are pursuing strategic communication as a focus are required to take the course from which the participant pool was heavily based. Those students, despite focusing on strategic communication, are still required to take multiple journalism-focused classes to fulfill their major requirements. In doing so, the participants from the Scripps school will have taken classes focused on journalism prior to the study. Those classes may have influenced how they view news/information programming. The strategic communication course does not have prerequisites and does attract students from outside of the journalism program, so those students will not necessarily have been exposed to the same in-depth analysis and understanding of the principles of journalism as those within the program. Therefore, future studies that involve judging media content from multiple genres (which also include news/information) and that also utilize students from multiple departments should consider control variables relating to the students’ major and background. It is unknown how many of the students involved in the study have taken other classes within the journalism department.

As mentioned in previous sections, along with their evaluations of the three shows, the participants also responded to the statement “I watch (program name) often” after watching each clip. The purpose of asking the participants to respond to that
statement was to understand the participants’ past experiences with the shows. However, the responses were unusual for a few reasons, which questions their validity. The descriptive statistics for the shows were as follows: *Better Off Ted* (M = 1.68, SD = .889), *Studio 60* (M = 2.25, SD = 1.252), and *TechKnow* (M = 1.98, SD = 1.174). First, the availability of the shows would lead one to believe that *TechKnow* and *Better Off Ted* should have higher previous experience than *Studio 60*. *Better Off Ted* aired for two seasons on ABC and is currently available on Netflix (where it is the fourth highest rated television show). While *TechKnow* is available to stream through Al Jazeera English’s website, the network was so poorly watched, averaging only 14,000 viewers for total day ratings (Chariton, 2014), that the network ceased broadcast at the end of April 2016 (Stelter & Kuldt, 2016).

In comparison, *Studio 60*, which only aired for a single season on NBC and is not available for streaming from any service online unless purchased on an episode-by-episode basis, had the greatest mean response to the statement. *Studio 60* was also the oldest show (airing almost 10 years prior to the experiment), which should indicate that participants would be less likely to be aware of it as it was the least readily available show. Given the inconsistencies, it appears that the participants may have misunderstood the question or given false statements. The wording of the question could be changed in future studies to help explain any issues within the current study. Inclusion of the word “often” in the phrasing may have confused participants. Additionally, because the questions were randomly ordered, the participants may have answered the question in
haste, without fully reading it, or confused it with the question on whether they would like to watch more of the show in the future.

From a methodological perspective, researchers are limited in the ecological validity that they can maintain for researching algorithms for a variety of reasons. Many algorithms utilized by websites and services are proprietary in nature and are unavailable to the public. Those algorithms may change without notice to users or researchers studying the algorithm’s effects. Researchers may create their own algorithms to test a variety of elements but cannot be sure if they are mimicking the effects that would be elicited from a specific website. Therefore, the findings from Study 2 are at least a starting point in understanding the overall influence of utilizing personalized content that can go into a larger research agenda focusing on the manipulation of algorithmic results and how users’ perceptions shift. Many of the technology companies that incorporate personalized content into their services perform research within their own corporate structure and examine the effectiveness of their builds on current users. By doing so, they may be the only ones who can most fully understand the outcomes they are creating by privately maintaining the records in such a way that the public cannot access them. In fact, they may not be examining the larger impact that they are having on their users, as it is unknown what they know and what they do not. After the company was accused of negatively influencing its users through its algorithm by helping the spread of “fake news” following the 2016 United States presidential election, Mark Zuckerberg, the founder and chief executive officer of Facebook, denied the proclamation that the website was manipulating its users without their knowledge (Wingfield, Isaac, & Benner, 2016).
However, at the same time, Facebook does proclaim that the network has been able to influence the users’ purchase intentions, as the social media site is heavily utilized as an advertising platform (Facebook, n.d.), yet somehow separates that from influencing political decision-making and knowledge.

There is an ethical argument to be made that the companies that utilize these algorithms should increase their transparency. If the algorithms have been shown to influence the users, the companies should not only release details on how the algorithms work but should also actively work to research the effects that they are having on their users. In addition, they should state in common terms how using the service may affect users and work to mitigate how others can take advantage of that service. For example, services that are mostly closed platforms like Netflix (where individuals are only able to rank media on a five-point scale) have less influence over others than more relatively open platforms like Facebook or Twitter. For the latter, the open platforms create a forum where the users are more vulnerable to outside influence from others who can manipulate the algorithm as they see fit.

In terms of limitations, the hypotheses are related to only enjoyment of content. More research will be needed to better understand non-video related content, other dependent variables (such as newsworthiness, information acquisition, and more), and other types of targeting. Future research may consider how different forms of recommendation systems make suggestions about the type of content a user would like, as opposed to simply giving a numerical rating on how the system assumes the user will evaluate the content. Future research could utilize demographic characteristics about the
user to give ratings. Research that surveys users on their beliefs and attitudes surrounding the presence of ratings/recommendations could help researchers better understand how personalization may influence the user.
GENERAL DISCUSSION

When taking both studies together along with the results of Saks (2013), a variety of implications and applications can be seen, as well as additional questions raised. For Study 1, the participants did not show a change in enjoyment when manipulating the types of commercials paired with content for individuals outside of their demographics, regardless of genre. Within Study 2, however, participants stated they enjoyed more the shows that were supposedly in line with their previous attitudes, as stated by the primed ratings. The general progression of logic from Study 1 to Study 2 focuses on how the participants felt about how content matches with their interests, whether inferred demographically (as is the case for Study 1) or explicitly stated based on supposed individual preferences (as is the case for Study 2). In Study 1 the participants were not effectively swayed by the inclusion of the commercials with the content, while participants in Study 2 were influenced by the primed ratings. The rest of this chapter will cover the various theoretical, methodological, and practical implications stemming from the findings, as well as limitations and suggestions for future research.

Theoretical Implications

Given that Study 1 and Study 2 utilized differing primary theoretical frameworks, it is difficult to judge the results of both studies in tandem. However, looking at the findings of both Saks (2013) and Study 1, it does seem that Mandler’s (1990) theory may not be the best fit moving forward in this line of research. In neither study did the participants register the necessary change in arousal required by the theory. This could account for the limited findings of only a statistically significant difference for the
explicit evaluation and approaching, but not meeting significance in the implicit evaluation in Saks (2013), while exhibiting essentially no significance in Study 1.

One could argue that the results from both studies, along with those from Saks (2013), demonstrate the importance of accurately demographically targeting the users/viewers. When the viewers saw content outside of their demographic, they did not seem to be as influenced as when the content was in their demographic. Additionally, there is potential for that demographic to be “falsified,” if one views the customized ratings as a means of triggering the elements in the viewers/users that would otherwise be alluded to with generalized demographic information. Although that is simply conjecture at the moment, it is something that should be further examined in light of the potential mechanisms underlying the mental processes.

The primary focus on the participants’ enjoyment of the media content, and the way in which the enjoyment response was manipulated, adds additional data for consideration in the task of creating a strong operational definition for media enjoyment. As mentioned in the Literature Review for Study 1, media enjoyment has been a highly utilized variable without a consistent operationalization. Further research is needed to fully understand the concept of media enjoyment, but these studies, particularly Study 2, can be included in data points to better understand it. Given the mixed results across all three dimensions of the media enjoyment concept for both studies, one must ask more about how each dimension comes together to create the entire concept. By having significance for one dimension and not the other two, do the participants experience some enjoyment or are all three parts required to affect the whole? Further quantitative and
qualitative research should examine the phenomena. Especially given that Expectancy-Disconfirmation theory has rarely been utilized with the concept of enjoyment, it is an area that should receive further study.

One could argue that the results from Study 2 were not from a difference in enjoyment due to ratings but simply from giving the participants a frame of reference upon which to judge the shows. However, the results found that in across five of the six outcomes for both the seven-point scale and indexed measures the control condition fell between the positive and negative conditions. One can assume that there was at least some additional processes in the minds of the participants other than simply comparing and contrasting their experiences to that of the initial ratings. The control group was not aware of the two ratings that the other groups received, so it is not possible for them to have intentionally or unintentionally placed their ratings between the other two, and only in one case, Better Off Ted, was change shown.

Methodological Implications

One element that differed between Study 1 and Study 2 is that the participants in Study 1 were not given any indication that the advertisements were different prior to seeing the content, nor did they know that the advertisements played any part in the outcome. Those participants in Study 2, however, were told prior to their involvement in the experiment that the purpose of the procedure was to analyze an algorithm that attempts to predict people’s media tastes based off of past experiences with other television shows. This difference between the two experiments could have caused participants to feel that stimulus materials varied in how well they connected with their
own personal preferences. An advertisement, paired with content, is subtly connecting the target audience to the intended demographic, while the demographics do not necessarily matter when the participants are told that the content rating is specifically personalized for them. By having the personalized rating based on their own preferences, the participants do not necessarily have to connect with overarching demographic characteristic. In the case of the personalized rating, however, they may not trust its accuracy.

One aspect that is open for various manipulations for the methodology, and outcome variables, is with regard to the type of experimental environment. For Study 1, the participants were watching the shows in a traditional media environment, on a television, with no control over the television, volume, etc.. Similarly, for Study 2, the participants were utilizing a controlled experimental program that does not reflect similar environments that one would experience in the real-world. While there are questions of ecological validity that arise from that (see Research Limitations and Future Studies section for further discussion of ecological validity in media effects studies), there are also important considerations that come from motivation within these different environments. If there participants do not feel adequately motivated to watch a show, either through their own interest or through additional means such as added control over the content, they may interact with the media differently and experience a different outcome. For example, a user on YouTube will have additional elements that are keeping their attention, including pop-up ads and video suggestions on the sidebar and those two
things may influence the outcome of enjoyment either positively or negatively. Therefore, motivation may shift and change the overall effect.

Particularly with the confusing results related to the participants’ past experiences with the shows in Study 2, it may be necessary to include more questions to assess the true nature of their experience with the show. That could be done in a few ways, one of which is pretest the content with separate participants. That way, instead of utilizing a “control condition,” during the experiment, per se, it could be possible to set a baseline measure for the participants’ responses prior to the experiment that may inform the questions and expected responses prior to the manipulations, if the participants for a larger study were first randomly assigned to condition and the target stimuli and questions counterbalanced, as described herein. Similarly, the commercials in Study 1 were pretested for the study for which they were used (Saks, 2013) to ensure that they were considered comparable in terms of visual and audio quality. That becomes more difficult when comparing other types of variables, such as enjoyment and relevance, across entire shows and genres. However, there are certainly some elements that could be used to help understand the full extent of the manipulations’ effects, including setting a wider range of available responses related to past experiences with a show. That might require that more qualitative responses be coded, such as “I’ve seen the entirety of the show before,” “I’ve seen a few episodes,” “I’ve never seen any episodes but I’ve heard of it,” or “I’ve never seen nor heard of the show before.”

In looking at similar studies, many utilized more variables in their understanding of the concepts being studied. One of the advantages of utilizing fewer measures and
analyses is that there is less likelihood of making a Type I error. However, it may be necessary to bulk up the measures to make sure that a Type II error is not committed by undermeasuring the concept.

Practical Implications

From a practical standpoint, one can see the influence of supposedly customized ratings and how they may be more effective than utilizing content based off of demographic cues. That thought should give pause to traditional media providers that may want to consider the ways in which their content can influence more personalized factors. As cable and satellite television providers are starting to include more “on-demand” offerings with their services, that is one area where the providers may be able to include recommendation features. Even without the type of measurement technology utilized by Nielsen to understand audience patterns, usage data from on-demand channels should be able to give more information on the likes and dislikes of users to allow for similar systems to the recommender systems utilized by websites. Such on-demand usage data could also be utilized to inform the type of commercial packaging decisions set by network executives. The technology available could allow on-demand offerings to switch out commercials for ones that are more targeted to specific viewers, much like what is done on YouTube and other websites that track user patterns to give more effectively targeted advertisements. There are certainly privacy and ethical considerations that the on-demand providers must consider, but those may not defer major corporations from considering such a proposal. Those concerns could potentially
be diminished if there is transparency in how the functionality works or by incentivizing the use.

Doing so may give users additional cause for concern over the use of targeting algorithms. Many may assume that they are being manipulated in ways they don’t quite know or understand, but any additional evidence that individuals’ perceptions can be influenced by forces with specific intentions or conflicts of interest may make users second-guess their use of those platforms. As mentioned in the Literature Review for Study 2, there are plenty of studies that have considered the ways in which individuals’ decision-making processes may be influenced by algorithms created by researchers, however if the simple act of making a randomly-assigned recommendation can also influence individuals, it should increase the average person’s skepticism toward the services.

Research Limitations and Future Studies

There are a few limitations to consider for both studies. As with almost all media-related experimental studies, there is a question of ecological validity and what that means when interpreting and understanding the results. Despite the fact that the idea of ecological validity is a highly-debated issue (Hammond, 1998; Reis & Gosling, 2010; Kvavilashvili & Ellis, 2004; Dobbins, Lane, & Steiner, 1988; Berkowitz & Donnerstein, 1982; Greenberg, 1987; Gordon, Slade, & Schmitt, 1986; Kardes, 1996), it is still something that one should mention when discussing laboratory studies given the misunderstanding around them. The concerns with the limitations of the experimental methodology (despite the fact that all methodologies have inherent limitations that cannot
fully encompass every aspect, detail, cause, effect, nuance, understanding, element, and so on) come from the participants in both studies being asked to come to a laboratory and watch media content outside the realm of their normal lives. It is not necessarily clear how they would have reacted to these manipulations at home, however we can say that in an experimental environment the shown effects did occur, as the ultimate goal is to understand the cause-and-effect relationship being controlling for as many factors as possible. For Study 1, the participants were in a simulated living room, where they sat on a couch while watching the content on a large television. At the same time, they were hooked up to the machine that registered their arousal levels with a researcher present in the back of the room. Those were trade-offs that were necessary given the limitations of the experiment, such as time, locations available, and monitoring the equipment to make sure it was functioning. Similar constraints were also present in Study 2, as participants needed to utilize a computer that would display the content and record their responses. Additionally, in the case of Study 2, there is little that can be done in the realm of ecological validity due to the nature of proprietary algorithms. Any study that focuses on the proprietary algorithms is only making a best guess as to how close their algorithm is to the one their algorithm is modeled after. With that in mind, the experiment was designed to bypass that by utilizing the backstory of testing an algorithm. Although the issue of ecological validity is still there, the participants were randomly assigned to the various conditions so that any individual level factors should even out across conditions. As such, the focus of the experiment was to understand the influence of making a recommendation/priming statement to the participants prior to watching the stimuli.
Therefore, one can see that the effects can still occur as a psychological response to the controlled factors. As such, that is in line with the goals and outcomes of experimental research in that the focus is on control of variables, cause preceding effect, and random assignment (Kardes, 1996). The research is intended to demonstrate an effect with a focus on internal validity instead of generalizability.

In multiple ways, there are opportunities for future research about a variety of elements seen within the two studies. Given that “enjoyment” was the dependent variable for both studies, utilizing a different focus may help researchers understand any potential phenomena that are occurring. The focus could include a different conceptualization of media enjoyment or an entirely different dependent variable, such as newsworthiness or relevance. A focus on newsworthiness could be enlightening within the context of both studies, as it would allow traditional media that utilize commercial breaks the opportunity to see if the ads influence how viewers consume the content. In the case of customized ratings, traditional media outlets could attempt to determine if individuals view “recommended” content differently.

In future studies research may also include the element of choice. Instead of guessing what type of content participants will like researchers may allow viewers to choose what to watch. That choice could either be paired with advertisements of varying demographic manipulations or with customized ratings. In giving the choice, it allows the participants to have a degree of control, which will eliminate uncertainty surrounding what they might like. While adding the element of choice eliminates random assignment, which is necessary to establishing a cause-and-effect relationship, as more online venues
allow users to select what to watch and when, instead of traditional media’s paradigm of scheduled viewing, there are additional effects that may come into play when the individuals feel like they are in control of what shows are given to them as opposed to having the information shown at them.

Building off of that, it would be valuable to know, from the user’s perspective, how viewers would react if they knew that they could be manipulated in terms of enjoyment. If they knew about that possibility, would they be more or less inclined (or not affected) to utilize those types of services? Would the effect still hold if the individuals knew of their potential to be influenced?

Conclusion

This dissertation focused on the ways in which participants’ perceptions of television show enjoyment could be influenced by external factors, including commercials and primed supposedly customized ratings. Those external factors were based on the inclusion of age-based demographic cues or a narrative explanation tying past viewing tastes to other shows. The target shows came from a variety of genres to examine the potential for enjoyment in different ways. The dissertation adds to the literature by building off of previous understandings of media enjoyment and theoretical knowledge.

The dissertation is but a first step in a potential research agenda as it creates multiple questions to be answered in future studies. Those questions, related to the influence of demographic targeting, algorithms, media enjoyment, and numerous other variables, should be at the forefront of future research.
REFERENCES


indicators. In *Proceedings of the 6th international conference on Intelligent user interfaces* (pp. 33-40). ACM.


Fazio, R. H. (1990). Multiple processes by which attitudes guide behavior: The MODE


not-so-minimal" consequences of television news programs. The American Political Science Review, 76(4), 848-858.


Stelter, B., & Kludt, T. (2016, March 27). The final days of Al Jazeera America. CNN


Wagner, C. B (2001). Implicit attitudes and anti-drug PSAs: Automatic processes and unreasoned action. Top Three Student Paper presented to the Communication Theory & Methodology Division at the 84th annual meeting of the Association for Education in Journalism and Mass Communication, Washington, DC.


APPENDIX A: EXAMPLE IMPLICIT ASSOCIATION TEST FOR STUDY 1

ID #: __________

PLEASE DO NOT OPEN UNTIL THE RESEARCHER ASKS YOU TO DO SO.

THANK YOU!!!
<table>
<thead>
<tr>
<th>THE GOLDEN GIRLS</th>
<th>COLORS</th>
</tr>
</thead>
<tbody>
<tr>
<td>MIAMI</td>
<td>BLUE</td>
</tr>
<tr>
<td>PROPOSAL</td>
<td>ORANGE</td>
</tr>
<tr>
<td>BLANCHE</td>
<td>PINK</td>
</tr>
<tr>
<td>FRIENDS</td>
<td>GREY</td>
</tr>
<tr>
<td>COMEDY</td>
<td>BROWN</td>
</tr>
<tr>
<td>SOPHIA</td>
<td>YELLOW</td>
</tr>
<tr>
<td>WIDOWS</td>
<td>PURPLE</td>
</tr>
<tr>
<td>TELEVISION</td>
<td>RED</td>
</tr>
</tbody>
</table>

STOP. DO NOT CONTINUE!
<table>
<thead>
<tr>
<th><strong>positive</strong></th>
<th><strong>negative</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>good</td>
<td>bad</td>
</tr>
<tr>
<td>pleasant</td>
<td>unpleasant</td>
</tr>
<tr>
<td>valuable</td>
<td>worthless</td>
</tr>
<tr>
<td>favorable</td>
<td>unfavorable</td>
</tr>
<tr>
<td>acceptable</td>
<td>unacceptable</td>
</tr>
<tr>
<td>nice</td>
<td>awful</td>
</tr>
<tr>
<td>wonderful</td>
<td>horrible</td>
</tr>
<tr>
<td>excellent</td>
<td>poor</td>
</tr>
</tbody>
</table>

**STOP. DO NOT CONTINUE!**
<table>
<thead>
<tr>
<th>COLORS</th>
<th>THE GOLDEN GIRLS</th>
</tr>
</thead>
<tbody>
<tr>
<td>RED</td>
<td></td>
</tr>
<tr>
<td>COMEDY</td>
<td></td>
</tr>
<tr>
<td>PURPLE</td>
<td></td>
</tr>
<tr>
<td>GREY</td>
<td></td>
</tr>
<tr>
<td>SOPHIA</td>
<td></td>
</tr>
<tr>
<td>PINK</td>
<td></td>
</tr>
<tr>
<td>MIAMI</td>
<td></td>
</tr>
<tr>
<td>WIDOWS</td>
<td></td>
</tr>
<tr>
<td>BLUE</td>
<td></td>
</tr>
<tr>
<td>PROPOSAL</td>
<td></td>
</tr>
<tr>
<td>BLANCHE</td>
<td></td>
</tr>
<tr>
<td>ORANGE</td>
<td></td>
</tr>
<tr>
<td>YELLOW</td>
<td></td>
</tr>
<tr>
<td>TELEVISION</td>
<td></td>
</tr>
<tr>
<td>BROWN</td>
<td></td>
</tr>
<tr>
<td>FRIENDS</td>
<td></td>
</tr>
</tbody>
</table>

STOP! DO NOT CONTINUE!
<table>
<thead>
<tr>
<th>positive</th>
<th>negative</th>
</tr>
</thead>
<tbody>
<tr>
<td>favorable</td>
<td>worthless</td>
</tr>
<tr>
<td>pleasant</td>
<td>unpleasant</td>
</tr>
<tr>
<td>unacceptable</td>
<td>valuable</td>
</tr>
<tr>
<td>wonderful</td>
<td>unfavorable</td>
</tr>
<tr>
<td>bad</td>
<td>excellent</td>
</tr>
<tr>
<td>unpleasant</td>
<td>awful</td>
</tr>
<tr>
<td>valuable</td>
<td>good</td>
</tr>
<tr>
<td>unfavorable</td>
<td>horrible</td>
</tr>
<tr>
<td>acceptable</td>
<td>nice</td>
</tr>
<tr>
<td>nice</td>
<td>poor</td>
</tr>
</tbody>
</table>

STOP! DO NOT CONTINUE!
COLORS or positive

FRIENDS
valuable
ORANGE
unacceptable
PINK
favorable
TELEVISION
worthless
BLUE
wonderful
PROPOSAL
horrible
SOPHIA
pleasant
BROWN
unpleasant
RED
awful
COMEDY
excellent
GREY
acceptable
WIDOWS
nice
MIAMI
bad
PURPLE
unfavorable
BLANCHE
good
YELLOW
poor

STOP. DO NOT CONTINUE!
COLORS or positive

unfavorable
BROWN
 excellent
BLANCHE
 bad
WIDOWS
 wonderful
YELLOW
 poor
PROPOSAL
 worthless
TELEVISION
 horrible
BLUE
 pleasant
RED
 acceptable
COMEDY
 unpleasant
SOPHIA
 good
GREY
 nice
PURPLE
 favorable
MIAMI
 unacceptable
FRIENDS
 awful
PINK
 valuable
ORANGE

STOP! DO NOT CONTINUE!
THE GOLDEN GIRLS

COLORS

COMEDY

GREY

TELEVISION

BLUE

PROPOSAL

BROWN

FRIENDS

ORANGE

MIAMI

PINK

BLANCHE

PURPLE

WIDOWS

YELLOW

SOPHIA

RED

STOP! DO NOT CONTINUE!
THE GOLDEN GIRLS or positive TELEVISION
good
RED
excellent
BLANCHE
horrible
PROPOSAL
valuable
YELLOW
poor
FRIENDS
nice
PINK
unpleasant
BROWN
favorable
WIDOWS
acceptable
BLUE
awful
SOPHIA
bad
GREY
wonderful
ORANGE
unfavorable
MIAMI
unacceptable
PURPLE
pleasant
COMEDY
worthless
COLORS or negative
STOP! DO NOT CONTINUE!
awful
FRIENDS
good
TELEVISION
acceptable
YELLOW
unfavorable
PROPOSAL
valuable
GREY
bad
SOPHIA
nice
RED
unpleasant
PINK
wonderful
COMEDY
horrible
BLUE
excellent
BROWN
poor
WIDOWS
pleasant
ORANGE
favorable
MIAMI
unacceptable
PURPLE
worthless
BLANCHE

STOP! DO NOT CONTINUE!
APPENDIX B: EXPLICIT EVALUATION FOR STUDY 1

Below is a list of word pairs. Circle one of the numbers near the word in each pair that best describes how you feel about the following statement:

“I think (name of show) is…”

1. bad 1 2 3 4 5 good
2. unpleasant 1 2 3 4 5 pleasant
3. worthless 1 2 3 4 5 valuable
4. unfavorable 1 2 3 4 5 favorable
5. unacceptable 1 2 3 4 5 acceptable
6. awful 1 2 3 4 5 nice
7. horrible 1 2 3 4 5 wonderful
8. poor 1 2 3 4 5 excellent
Overall, I liked the television program I just watched

<table>
<thead>
<tr>
<th>Strongly Disagree</th>
<th>Disagree</th>
<th>Neutral</th>
<th>Agree</th>
<th>Strongly Agree</th>
</tr>
</thead>
</table>

Overall, I liked the television program I just watched

<table>
<thead>
<tr>
<th>Strongly Disagree</th>
<th>Disagree</th>
<th>Neutral</th>
<th>Agree</th>
<th>Strongly Agree</th>
</tr>
</thead>
</table>

I watch (name of show) often

<table>
<thead>
<tr>
<th>Strongly Disagree</th>
<th>Disagree</th>
<th>Neutral</th>
<th>Agree</th>
<th>Strongly Agree</th>
</tr>
</thead>
</table>

I would like to see more episodes of (name of show) in the future

<table>
<thead>
<tr>
<th>Strongly Disagree</th>
<th>Disagree</th>
<th>Neutral</th>
<th>Agree</th>
<th>Strongly Agree</th>
</tr>
</thead>
</table>
APPENDIX C: EXPLICIT EVALUATION FOR STUDY 2

Below is a list of word pairs. Circle one of the numbers near the word in each pair that best describes how you feel about the following statement:

“I think (name of television show) is…”

1. bad 1 2 3 4 5 good
2. unpleasant 1 2 3 4 5 pleasant
3. worthless 1 2 3 4 5 valuable
4. unfavorable 1 2 3 4 5 favorable
5. unacceptable 1 2 3 4 5 acceptable
6. awful 1 2 3 4 5 nice
7. horrible 1 2 3 4 5 wonderful
8. poor 1 2 3 4 5 excellent
APPENDIX D: ADDITIONAL QUESTIONS FOR STUDY 2

Overall, I liked the television program I just watched

<table>
<thead>
<tr>
<th>Strongly Disagree</th>
<th>Disagree</th>
<th>Neutral</th>
<th>Agree</th>
<th>Strongly Agree</th>
</tr>
</thead>
</table>

I watch (name of television program) often

<table>
<thead>
<tr>
<th>Strongly Disagree</th>
<th>Disagree</th>
<th>Neutral</th>
<th>Agree</th>
<th>Strongly Agree</th>
</tr>
</thead>
</table>

I would like to see more episodes of (name of television show) in the future

<table>
<thead>
<tr>
<th>Strongly Disagree</th>
<th>Disagree</th>
<th>Neutral</th>
<th>Agree</th>
<th>Strongly Agree</th>
</tr>
</thead>
</table>