Do We Judge Fiction by the Author's Gender?

Ivanski, Chantelle; Humphries, Stacey; van Dalen-Oskam, K.H.; Mar, Raymond A.

published in
Journal of Media Psychology
2022

document version
Peer reviewed version

document license
Unspecified

Do we judge fiction by the author's gender?

Chantelle Ivanski[1], Stacey Humphries[2], Karina van Dalen-Oskam[3], & Raymond A. Mar[1]

[1]York University, CA

[2]University of Pennsylvania, USA

[3]Huygens Institute, NL

Author's Note:

Please direct correspondence to C. Ivanksi.

York University, Department of Psychology

4700 Keele St. W., Toronto, ON, Canada. M3J1P3

civanski@yorku.ca

416-736-2100 x20769

Abstract

Female authors of fiction often perceive themselves to be undervalued in relation to their male counterparts. What is not clear is whether this preference for male authors comes from readers or publishers. Two pre-registered studies examined how university students evaluated book passages attributed to either male or female authors, and investigated whether negative evaluations of romance novels are based on their association with women. In Study 1, participants read identical passages attributed to either male or female authors and evaluated them. Study 2 extended this work by adding attributions of genre: either romance or literary fiction. Linear mixed-effects modeling and Bayesian analyses were employed to analyze these data. Study 1 demonstrated little preference for books attributed to males over females and Bayesian analyses confirmed support for the null in most cases. The results of Study 2 similarly suggested that author gender and genre attributions do not have a strong influence on evaluations.

Keywords: Narrative Fiction, Gender Stereotypes, Reading, Discourse, Evaluations

Introduction

The question of whether writing is judged more negatively when attributed to women relative to men has been investigated extensively, in both academia and workplace contexts (Gallivan, 1991; Lebuda & Karwowski, 2013; Moore & Trahan, 1998). However, one domain in which there has been little research on this topic is literary fiction. This is somewhat surprising, as female authors of fiction often perceive themselves to be underrepresented and undervalued (e.g., Flood, 2015). Recent research on Dutch readers of fiction found that fiction novels written by female authors were perceived as less valued for their literary quality than novels written by male authors. This same research also found that there was no gender difference in perceived literariness for suspense novels, and that romance novels were valued the least in terms of literary quality (Koolen, 2018; Koolen et al., 2020). Within the realm of fiction, the romance genre is rather unique, as it is largely by and for women and boasts immense popularity yet is almost universally viewed as being of poor quality. To build on this research from the Netherlands, we investigated whether fiction attributed to women is judged more harshly than that attributed to men in a Canadian context, and whether it is the romance genre's connection to women that explains its poor reputation. To investigate these questions, we examined whether participants would regard identical pieces of fiction as being worse in quality when they are attributed to women compared to when they are attributed to men (Study 1). We also examined if participants would differ in their evaluations when they were told the passages were from a romance novel rather than a work of literary fiction (Study 2)

Judging Writing based on the Author's Gender

Research on how writing by males and females is evaluated began with a seminal study by Goldberg (1968). In this experiment, female participants first rated occupations as male,

female, or neutral. Next, a second set of female participants rated articles from each of the fields. Although all participants read the same texts, they were told the articles were written by either a male or a female. In contrast to expectations, it was found that male authors were rated as more competent and having written more valuable articles than female authors across all six fields (Goldberg, 1968).

Attempts to replicate what was dubbed 'The Goldberg Effect' have been plentiful (i.e., women judging women authors more negatively than men), with some using just female participants, as in the original study, and some including male participants. However, the results of these studies have been mixed. Some have found the same bias towards rating male authors as superior (e.g., Gallivan, 1991; Lebuda & Karwowski, 2013; Moore & Trahan, 1998; Paludi & Bauer, 1983), others only find more negative evaluations of females for articles about stereotypically-male subjects (e.g., Haemmerlie & Montgomery, 1991; Mischel, 1974), and some find no bias based on gender at all (e.g., Levenson et al., 1975; Pirri et al., 1995; Zhang et al., 2009). In an attempt to investigate whether an overall "Goldberg Effect" exists, a meta-analysis on these studies was conducted (Swim et al., 1989). Based on 123 studies from 106 articles, there appears to be no consistent evidence of a difference in evaluations based on the gender of the author (Swim et al., 1989). However, effects seem more likely to emerge when little information about the author is presented aside from gender. Likely as a result of the many null findings, research on this topic peaked in the 1970s and is no longer commonplace (Pirri et al., 1995).

Despite this extensive research on professional and academic writing, very little research has looked at the evaluation of artistic or creative products, although there are some exceptions. For example, an experiment examining ratings of New Age music found that male composers were rated more positively than female composers (Colleyet al., 2003). In another experiment on

painting, female painters with common names (e.g., Anna) were rated lower in ability than males, anonymous artists, and females with unique names (e.g., Lea) (Lebuda & Karwowski, 2013). This bias against females with common names also extended to musical compositions, but not poetry, with female poets with common and unique names rated similarly to males (Lebuda & Karwowski, 2013).

With respect to research on fiction, however, we are aware of only one prior experiment along these lines (Ciechanowicz, 1983), in addition to the Dutch findings that women authors are less valued for their literary contributions (Koolen, 2018; Koolen et al., 2020). In this experiment, 450 participants read one of three texts, one of which was a fictional narrative that told the story of a man trying to regain the love of his wife. Participants were told the author was either a male, a female, or were asked to guess the author's gender. They were then asked to evaluate the author. When authorship was attributed to a female, the author was rated more positively on all dimensions than when the text was attributed to a male (i.e., more intelligent, credible, nice, and sophisticated). When they were not told the gender of the author, however, participants tended to guess that it was written by a male, but this did not affect how they rated the article (Ciechanowicz, 1983).

Although there is a paucity of empirical research on whether fiction authors are evaluated based on their gender, particularly experimental work, there is plenty of anecdotal evidence of a bias favoring men. For example, there is a long history of women writing under male pseudonyms, or using gender-ambiguous names, to avoid a perceived prejudice against women in publishing. This includes high profile examples within classic literature (e.g., Mary Anne Evans writing *Middlemarch* under the pseudonym George Elliot), but also modern day examples (e.g., Joanne Rowling writing the Harry Potter series as J.K. Rowling) (Armitage, 2018). This

parallels how ethnic minorities tend to change their names on resumes to be more "white," often improving their job prospects(Kang et al., 2016). In a similar vein, one female author claimed that she was eight times more likely to be published when using a male pseudonym (Denham, 2015). In addition to these anecdotes, there is also some objective evidence of bias against female authors. For example, books written by female authors are sold for less than those authored by males (Weinberg & Kapelner, 2018). Most of this negative bias, however, seems to stem from the publishers who select authors and set prices, rather than from readers. It is not clear whether any bias that may exist in publishers is a reflection of actual reader preferences. An example of this asymmetry can be seen in film. Historically, films with female protagonists have been viewed as unprofitable by film companies based on the belief that teenage boys make up the majority of movie sales and would view these movies negatively (Prasad, 2018). However, in the past few years, movies with female leads have been extremely profitable, and were actually the top grossing films of 2017 (Owen, 2018). In the case of literary fiction, there is one genre that does appear to be viewed negatively by most readers: Romance novels. Because romance novels are also closely associated with women (as readers and authors), incorporating an examination of this genre may add additional nuance to the question of how fiction is perceived based on author gender.

Negative Evaluations of Romance Novels

Romance novels initially grew in popularity in 1970, when the publisher Harlequin began circulating them in America (Brackett, 2000). Since then, romance has grown to be the second-most commonly-read fiction genre (with mystery/thriller/crime being the most popular) (Statisitca, 2015). According to the Romance Writers of America Association (n.d.), in order to be considered a romance novel two criteria must be met: (1) there must be a 'central love story'

(i.e., the main plot of the book should be a relationship between two people), and (2) the ending must be optimistic and satisfying to the reader (i.e., a happy ending). Readers of romance novels are 82% female and mostly between the ages of 25 and 34 years-old (Romance Writers of America, n.d.). With respect to the authors, one estimate found that 99% of romance authors are female (Lois & Gregson, 2015). However, this estimate may be slightly inflated. In contrast to common practice for other genres, some male authors of romance write under female pseudonyms (Bookish, 2014). Male and female writers both experience stigma for writing romance, although the stigma is of different kinds. Female writers are shamed for openly discussing sexuality, whereas males are scrutinized for entering into a feminine sphere (Lois & Gregson, 2015). Thus, the criticism that male authors are subjected to is based on their perceived proximity to women and thus an association with femininity.

Despite the popularity of romance novels, there is widespread shame associated with reading books from this genre. Romance novels are often referred to as being "smutty" or "trashy," or as "porn for women" (Lois & Gregson, 2015). One romance publisher actually provides a free dust jacket to help readers hide the fact that they are reading a romance novel while in public (Brackett, 2000). The increasing availability of e-readers, where the book being read cannot be seen by onlookers, has led to an increase in the sale of romance novels (Akbar, 2012). Today, about 50% of all romance novels sold are for e-readers, compared to 20% for general fiction (Akbar, 2012). Romance readers themselves also criticize the genre for being simple and lacking in quality, and even mock other romance readers. In one study of romance readers, all of the participants said some version of "I'm not like most romance readers" (Brackett, 2000).

This highlights an interesting contradiction. Despite being the second-most commonly read genre of fiction, there is an almost universal condemnation of romance novels, even among those who read them. Given that romance novels are widely viewed as being by and for women, the question arises of whether romance novels are so disliked in part because of their association with women. This may parallel, in some ways, a possible belief among publishers that books by women are of less interest to consumers.

The Current Research

The first goal of the current research is to add to the body of research investigating whether fiction texts are evaluated more harshly when attributed to female authors compared to male authors. The second goal is to determine if the negative evaluations of romance novels can be attributed, in part or whole, to their association with female authors. To pursue these goals, two pre-registered studies were conducted. In Study 1, participants read four passages, two supposedly written by male authors and two by female authors, with no genre information provided. The gender of the author attributed to each passage was counterbalanced across participants, but the content of the passages was identical across participants. Participants then rated the quality of the passages and reported their enjoyment. Given the mixed results for this topic, this study was exploratory. In the absence of a strong directional prediction, the goal of this study was to help uncover whether any readers rate texts differently depending on the gender of the author attributed to the text.

Study 2 was identical to Study 1, with the addition of book genre as an additional factor. Participants read four passages purported to be from either romance novels or works of literary fiction. Literary fiction was chosen as a comparison group because it is a genre that is both held in high esteem and male-dominated. As an example of this dominance, women have received

only 14 of the 115 Nobel Prizes in Literature (Nobel Media, 2019). Further, the aforementioned Dutch research estimated that the percentage of female authors in the Dutch field of literary fiction is between 35 and 40%, but that the percentage of female authors receiving a literary prize is between 21 and 25% (Koolen, 2018). For this study, we anticipated two possible patterns of results. First, excerpts attributed to literary fiction might be rated higher than those attributed to romance novels, with little difference caused by the gender of the attributed author. In this case, excerpts purportedly from romance novels are expected to be viewed less positively than those for literary fiction. This pattern of results would suggest that romance novels are viewed as intrinsically bad, but not because they are written by women. The second possible pattern is that male authors would be viewed more positively than female authors for both literary fiction and romance. This would suggest that it is not romance that is judged to be of low quality, per se, but rather it is writing by women that is viewed more negatively. All methods, hypotheses, and proposed statistical analyses for both studies were preregistered on aspredicted.org, prior to the data being analyzed: https://osf.io/vj2pb/?view_only=0343f5f313914933862cc2ab5f15567f.

Study 1 Method

Participants

To account for potential exclusions (see criteria below), we planned to recruit 350 participants but ended up recruiting 427 psychology undergraduates because the recruitment platform allows people to continue to sign up until the number of completed surveys meets the criterion. Thus, participants were permitted to start the study even though more than 350 people had begun the process of completing it. Of these 427 participants, 230 of whom completed the study in lab (197 completed it online)[1]. Participants were recruited from a large Canadian university and were compensated with course credit.

Prior to analyzing the data, but during collection, we preregistered our exclusion criteria, stating that individuals would be excluded if they were under 16; failed 2 or more of the 8 comprehension questions, or both questions for a single passage; if they had read the passage previously; or if they guessed the purpose of the study. Upon exploring the data, it became apparent that the exclusion criteria regarding the comprehension items were exceedingly strict. Specifically, excluding participants who failed to answer two or more questions correctly resulted in a greatly reduced sample size of $N = 166$ (one person was also removed for guessing the purpose of the study). In addition to lowering our power, it is possible that participants who were responding conscientiously were excluded, simply because the comprehension questions were too difficult. Thus, these data were also reanalyzed with a new exclusion criterion, removing participants who scored one standard deviation below the mean for the comprehension questions. This change resulted in the removal of participants who answered less than 5 of the 8 comprehension questions correctly. All other exclusion criteria were left unchanged, and this resulted in a sample size of 290 participants. In order to maintain transparency, results based on the altered exclusion criteria are available on our OSF page (https://osf.io/vj2pb/?view_only=0343f5f313914933862cc2ab5f15567f). However, here we report results based on the original, pre-registered exclusion criteria ($N = 166$), since the results did not depart substantially when using a more liberal exclusion criteria. Most participants in this study were female (69.88%), ranging in age from 17 to 31 ($M = 19.54$, $SD = 2.15$). Detailed demographics, as well as all measures and materials, can be found on OSF.

Measures

Author names. Participants were shown 4 passages, each associated with an author name (2 male and 2 female), based on 12 possible first names. These names acted as paratextual

information communicating the author's gender. The names were pretested in a pilot study prior to data collection, to ensure that they were not strongly associated with constructs relevant to the evaluations (e.g., intelligence, education, creativity; $N = 165$, pilot results on OSF). Last names were the same for both male and female authors, selected at random from common US surnames. Names were presented with a short biography that was identical for both female and male authors, preceding the presentation of the passages. Biographies were included to further reinforce the author's gender by using 'he' and 'she,' throughout.

Passages and comprehension questions. Participants read and evaluated 4 of 12 possible passages. Passages were excerpted from real works of fiction and ranged in length from 240–347 words. Each passage was presented with a fake title, and one of the author names and biographies. We modified each passage to include a grammatical error, to avoid floor effects when asking participants about any potential errors in the writing. After reading the passages, participants completed two comprehension questions, with the author's name presented once again to reinforce the author's gender. Comprehension questions were included in lieu of a direct manipulation check, because directly asking participants about the gender of the author would reveal the purpose of the study. Instead, comprehension questions were employed to measure whether participants attended to the text and were likely to have encoded author gender.

Evaluations. As our key dependent measure, participants evaluated the passages based on 6 questions[2], with responses provided on a 5-point Likert scale (*Strongly Disagree* to *Strongly Agree*). These items were: "I found this passage enjoyable to read," "I would be interested in reading more by this author," "I thought this passage was well-written," "I noticed grammatical errors in this passage," "I thought the characters were realistic," and "I thought that emotion was conveyed in this passage."[3]

Procedure

Participants were told that they would be shown a series of passages from books and asked questions about these passages, in order to better understand people's preferences in reading fiction. Next, they were shown a book title with one of the author names and biographies discussed above. Following this, participants were shown the first passage, selected at random. The passages were presented with male and female author names appearing with equal frequency, both within participants (2 male authors, and 2 female authors) and across participants (each passage was associated with both a male and female author with equivalent frequency). After reading the passage, participants completed the comprehension and evaluation questions. This procedure was repeated for the remaining passages.

Study 1 Results

Evaluations of the passages were analyzed using linear mixed-effects (LME) modelling with the LME4 package (Bates, et al., 2015) and the lmertest package (Kuznetsova et al., 2017) in R (R Core Team, 2014). Individual models were fit for each of the evaluation items. Random intercepts for participants and passages were included. It was not possible to include random slopes as these models failed to converge. Each evaluation question was analyzed as a separate dependent variable. We had also preregistered that we would include a total evaluation score by aggregating all the evaluation items, however, after conducting a confirmatory factor analysis it was found that the model had poor fit and will, therefore, not be included in the main paper. For transparency, however, we have still included these results in the supplemental materials. For our models, we first estimated a model with author gender (male vs. female) as a within-subjects design effect (Model 1). Next, a similar model was estimated with participant gender included as

an additional fixed-effect along with author gender (male vs. female) (Model 2). The fixed effects were allowed to interact for all models. Detailed frequentist results are available on OSF.

Models that included only author gender (excluding participant gender) were all statistically non-significant. That is, there were no major differences, based on gender of the author attributed to the text, in how much participants liked a passage (β = -.10, p = .15), whether participants thought the passages were well-written (β = -.01, p = .81), thought the characters were realistic (β = -.01, p = .92), thought the passages showed emotion (β = .01, p = .85), wanted to read more by the author (β = -.05, p = .51), or noticed errors in the passages (β = -.12, p = .06).

For models including both author gender and participant gender, there was a very small main effect of author gender, with participants reporting more liking for the passage when it was attributed to a male author relative to a female one (β = -.28, p = .035), however this did not survive a test of multiple comparisons based on marginal means. There was no main effect of participant gender (β = .02, p = .88) and no statistically significant interaction between author gender and participant gender for liking (β = .25, p = .11). The Akaike information criterion (AIC) was practically identical for the liking model that only included author gender (1883.7) and the model that included both author and participant gender (1883.3), suggesting that adding participant gender to the model did not much improve model fit. In fact, neither model was preferred over a null model including only the random effects (AIC = 1883.8).

We also found a small interaction between author gender and participant gender (β = .24, p = .05) on thinking characters were realistic. Male participants thought the characters were more realistic when written by a male author and female participants thought the characters were more realistic when written by a female author. Importantly, however, the effect sizes in these groups

are small, with mean differences of less than 0.5 between them and this did not survive a correction for multiple comparisons. Further, AIC was identical (1562.2) for the model that only included author gender and the model that included both author and participant gender, suggesting no improvement in fit with the inclusion of participant gender. In addition, neither model was preferred over a null model (AIC = 1560.2).

Lastly, we did not observe any main effects or interactions for the rest of our evaluation items, including whether participants noticed errors, thought the passages were well-written, wanted to read more by the authors, or thought that the passages showed emotion. Results for female participants are shown in Figure 1 and for male participants in Figure 2.

Our analysis treated the 5-point Likert items as continuous variables. However, linear mixed-effects models may be inappropriate for ordinal data (Bauer & Sterba, 2011). We therefore reanalysed the data using the "ordinal" package in R to estimate cumulative link mixed models (Christensen, 2019), but the results did not differ (code to reproduce these analyses are in our posted script on OSF).

Bayesian Analysis

The majority of our comparisons were not statistically significant. However, under a null-hypothesis significance testing framework we cannot assess the likelihood of the null hypothesis (i.e., that there is no difference in evaluation based on attributions of author gender). To evaluate evidence in favour of the null we conducted an exploratory analysis using Bayes factors that was not pre-registered. These analyses were calculated using the BayesFactor package in R to compute Bayes factors for mixed models (Morey et al., 2018). All analyses set a weakly informative prior on each fixed effect, represented by a Cauchy distribution with a scale factor of $r = .5$. This scale parameter of 0.5 corresponds to a probability of 80% that the standardized

effect-size of each fixed effect lies between -1.54 and 1.54. The scale parameter for the random effects in all analyses was set to $r = 1$, which means there is an 80% chance that the standardized effect-size of each random effect lies between -3.08 and 3.08. As above, random intercepts were included for participants and passages. Bayes factors were computed comparing models containing the random effects and various combinations of the fixed effects against a null model containing only the random effects. Bayes factors closer to 0 are considered as evidence in favor of the null, with values less than .3 considered substantial evidence for the null. In comparison, Bayes factors around 1 are considered inconclusive with values above 1 seen as evidence in favor of the alternative hypothesis, and values above 3 constituting good evidence. In the present study, Bayes factors indicated substantial evidence in favour of the null hypothesis in almost all cases (Table 1). One exception is whether participants wanted to read more from the author, with this Bayes factor closer to 1 (i.e., .93), indicating no clear evidence. Note that the results of this analysis at times contradict those found using the LME models. We report above a main effect of author gender for liking ($p = .035$; $BF = .25$), and an interaction between author gender and participant gender for character realism ($p = .053$; $BF = .01$), but in both cases the Bayesian results indicate evidence in favour of the null.

Study 1 Discussion

In Study 1 we found little difference in how respondents evaluate texts attributed to male or female authors. Some small, marginally statistically significant main effects were found, showing that participants liked texts better when they were attributed to a male author rather than female author, regardless of their own gender. This was observed for the individual item on liking. These results did not survive a correction for multiple comparisons, however. In addition, Bayes factors calculated for these same effects indicated that there was substantial evidence

favouring the null hypothesis. Thus, although there is evidence that publishers charge more for books written by male authors (Weinberg & Kapelner, 2018) or choose to publish more works with male (or gender-neutral) names (Denham, 2015), the current research suggests this bias is not mirrored on the part of the reader when evaluated text excerpts.

Using frequentist statistics, we also found a small interaction in which male participants thought characters were slightly more realistic when attributed to male authors, with female participants thinking the same for female authors. However, this effect did not survive a correction for multiple comparisons and Bayes factors for the same effect again contradicted this conclusion, indicating substantial evidence in favour of the null.

Finally, we found that author gender did not influence the other evaluation criteria (noticing errors, writing quality, and amount of emotion). Taken together, our results suggest that there was little evidence for, and often substantial evidence against, the presence of any reader bias based on the author's gender. In Study 2, we further explored whether attributions of author gender influence reader evaluations by including book genre as an additional variable: whether the excerpt was attributed to romance or literary fiction.

Study 2

Study 2 examines whether the negative evaluations of romance novels are due to the genre itself or the gender of the author. Participants evaluated passages supposedly taken from either romance novels or works of literary fiction, that were also ostensibly written by either a male or female author. In this study, there were two competing hypotheses: (1) that male authors would be rated more positively than female authors, regardless of genre, suggesting that the derogation of romance novels stems from its association with women; or (2) that romance would be rated worse than literary fiction regardless of author gender, suggesting that romance novels

are seen as intrinsically of poorer quality, regardless of author gender. This study mirrored closely the design of Study 1, aside from the added inclusion of a genre label for the excerpt.

Study 2 Method

Participants

To account for potential exclusions, we recruited 330 undergraduate students registered in a first-year psychology course at a large Canadian university. A total of 163 participants completed the study online and 167 completed it in-person, and all participants were compensated with course credit. The same data exclusions as in Study 1 were preregistered for Study 2. As a result, the same concerns regarding the strictness of the exclusion criteria for the comprehension items arose in Study 2. Following the original exclusion criteria resulted in a greatly reduced sample of 128 participants. Thus, an identical procedure as for Study 1 was followed, such that alternative criteria were also adopted excluding participants who scored one standard deviation below the mean or less (corresponding to answering less than four questions correctly). Here we report the results using the original exclusion criteria; the results from the analyses using this alternative exclusion criteria can be found on OSF as they did not differ substantially from the results obtained using the original exclusions. Participants were mostly female (68.75%, $n = 88$), with an age range of 18−56 years ($M = 19.94$, $SD = 4.18$).

Measures

Measures were identical to those used in Study 1, with the addition of genre information for each passage. Specifically, the biography presented with each author now included whether the author writes romance or literary fiction. Further, before seeing any names or passages, participants were told that they would be reading passages from literary fiction or romance novels. Study 2 was a split-plot design in that all participants saw two passages ostensibly

written by male authors and two by female authors; however, each participant saw only one purported genre. Book genre was included as a between-subjects variable so that participants would have less cognitive strain and to strengthen the manipulation. Doing so meant readers did not have to remember what genre they were reading, in addition to processing author gender. It also allowed us to reduce the study time to help prevent fatigue. Otherwise, the procedure for Study 2 mirrored that for Study 1. As with Study 1, all associated measures and materials for Study 2 can be found on OSF.

Study 2 Results

The analyses for Study 2 mirrored those performed for Study 1, with evaluations analyzed separately, with individual LME models for each of the evaluation items. Similar to Study 1, we had preregistered that we would examine an aggregate total measure, however, again the model fit was poor and is therefore not included in the results. Again, for transparency, we have still included these results in the supplemental materials. To compare the results with Study 1, we recreated the models with just author gender and participant gender as fixed effects (Model 1). The next set of analyses examined the effect of book genre by itself (romance vs. literary fiction), as a between-subjects fixed-effects variable (Model 2). Following this, additional models were evaluated with author gender (male vs. female) and book genre included (Model 3). The final analysis included all three variables: participant gender, author gender, and book genre as fixed effects (Model 4). In all analyses, the fixed effects were allowed to interact, and random intercepts were included for both participants and passages. As with Study 1, we will present frequentist results online in the supplementary materials and Bayes factors in the manuscript for all models.

In the first models recreating the results of Study 1, there were no statistically significant effects of author gender or participant gender on any of the evaluation items (i.e., liking, noticing errors, whether the passages were well-written, realism of characters, whether the passages showed emotion, wanting to read more by the author). In the models with only book genre as a fixed effect, there were no statistically significant effects of genre on any of the variables of interest. When models included both author gender and book genre as fixed effects, there were also no main effects or interactions for any of the ratings. Finally, when book genre, author gender, and participant gender were all included as fixed effects, we observed no statistically significant main effects on the variables of interest. Additionally, there were no statistically significant interaction effects for whether participants liked the passage (β = -.12, p = .73), noticed errors (β = -.08, p = .80), thought the passage was well written (β = -.19, p = .56), thought the characters were realistic (β = -.25, p = .33), thought the passage showed emotion (β = .30, p = .35), or wanted to read more by the author (β = .31, p = .38). Results for female participants are shown in Figure 3 and for male participants in Figure 4. As in Study 1, the results did not differ when cumulative link mixed models were estimated and the data were treated as ordinal (see scripts on OSF).

Bayesian Analysis

As in Study 1, we completed an exploratory analysis by calculating Bayes factors for the influence of author gender, participant gender, and book genre on each rating. In most cases, the Bayes factors indicated strong evidence in favour of the null hypothesis. Inconclusive Bayes factors were found for the effect of participant gender on noticing errors in the writing, and for the effect of participant gender on wanting to read more from the author. In no case did any Bayes factor indicate evidence in favour of the alternative hypothesis. With the addition of book

genre, the number of potential combinations of main effects and interactions became quite large (18 potential models for each dependent variable). For simplicity, we report the Bayes factors associated with each main effect separately, and the full model which includes all possible main effects and interactions (see Table 2).

Study 2 Discussion

This study tested two competing hypotheses. The first was that male authors would be rated higher than female authors regardless of genre, suggesting that the derogation of romance novels is due to its association with women. Second, that romance novels would be rated lower than literary fiction novels regardless of author attribution, suggesting that romance novels are inherently of poorer quality. Overall neither of these hypotheses were supported. The results of this study suggest that the respondents did not differ in their evaluations of fiction excerpts based on whether it was attributed to either male or female authors, or the romance or literary fiction genres.

The lack of differences in evaluation based on genre observed here is interesting given the seemingly universal low regard for romance novels. If participants judge excerpts attributed to romance as equivalent to those attributed to literary fiction when the text is identical, then it appears the mere label of "romance" is not enough to lead participants to believe the writing must be poor. During debriefing, multiple participants mentioned that they did not think the passages were particularly romantic given that they were ostensibly from romance novels. In truth, only one of the passages was actually taken from a romance novel. Because the main goal of this study was to examine the influence of paratextual information (i.e., attributions of genre and author gender), it was not our aim to have participants evaluate actual passages from romance novels. Thus, participant judgments seem to have been based primarily on the actual

content of the passage, rather than the genre to which they were attributed. It may be that romance novels are inherently of lower quality than other genres and, had more passage from romance novels been included, those would have been rated lower regardless of the attributed genre. In addition, rather than the writing, it may be the actual subject matter or content of romance novels that people find objectionable. Overall, the results from this study were not as we predicted, but they do offer some interesting insight into how people evaluate fictional texts and romance novels.

General Discussion

Study 1 found little difference in the evaluation of texts attributed to male or female authors, with Bayes factors indicating evidence in favour of the null hypothesis in most cases. In Study 2, no differences in evaluation were observed based on attributed author gender or genre. Bayes factors provide a valuable sensitivity check when frequentist results are not statistically significant to help determine whether there is evidence in favor of the null (Dienes, 2014). In this case, Bayes factors revealed that there was substantial evidence in favour of the null hypothesis for most evaluation criteria, or inconclusive evidence. These results are similar to those reported in the meta-analysis of the Goldberg effect, which found no differences in ratings of writing quality based on author gender, but for academic and professional work (Swim et al., 1989).

Our results suggest that there is very little evidence of bias against female authors by university students. These results pose an interesting question: if people do not differentiate between male or female authors, why do publishers show bias towards male authors (Denham, 2015)? And why are books written by women sold for less than those written by men (Weinberg & Kapelner, 2018)? More research will need to be done in order to understand why authors feel there is such a discrepancy in publishing between male and female authors. It is possible that a

gender bias does exist, albeit one perpetuated exclusively by publishers. If publishers believe that readers are less likely to purchase, or to pay less for, books by female authors, they may be disincentivized to publish those books. In doing so, they are enacting a gender bias that does not accurately reflect the opinions of consumers. Historically, the senior positions within publishing have been held by men (Milliot, 2019). If it is the case that these male executives believe that books by women are less likely to be successful, the results of the current study suggest these assumptions may be outdated. This male dominance in publishing is changing, however, with more women working in publishing than men and, since 2017, women holding more senior management positions than men (Milliot, 2019). It is also possible that biases and prejudices against female authors exist when evaluations are based solely on the book's cover (which contains the authors name), as opposed to when actual text excerpts are being evaluated as in these studies.

The similar evaluations for excerpts attributed to either romance or literary fiction suggest that there is something besides the label of 'romance' that drives negative judgements of books from this genre (Brackett, 2000). It may be something about the way romance novels are written that make them objectively worse than other forms of fiction, despite their popularity. One parallel is the phenomenon of reality TV. Reality TV shows are directly marketed to women, and are viewed the least favourably relative to other genres, and yet also remain incredibly popular (Shevenock, 2018). This again raises the question of why media that is targeted to women is so popular, but also typically viewed so negatively.

Limitations and Future Directions

A limitation of the current research was the relatively small sample size analyzed. These small samples may have made it difficult to detect small effects, small effects that, when

aggregated at the population level, would still result in a noticeable bias against female authors. These small samples resulted from our pre-registered exclusion criteria. But it is not clear whether the participants who were removed due to these criteria were less conscientious or if the comprehension questions were too difficult. It may be that the comprehension questions worked as intended and those who met the preregistered exclusion criteria were also more likely to have attended to and encoded the gender of the author, making our analysis of this smaller subset appropriate. In any case, re-analyzing our data with a more liberal exclusion criteria did not change our main results. In addition to the small sample size, it is important to note that this research was conducted using university students. Our results are therefore unlikely to generalize to the general population, but instead likely represent the behaviors of adults of similar age and background. Future research is necessary to see if these results are also observed for older adults.

One other possible issue is that the presentation of texts that lacked romantic content for Study 2 might have resulted in participants not believing that the passages were from actual romance novels. If that is the case, then this may have weakened our experimental manipulation. A future study could use more excerpts from actual romance novels to identify the specific factors that contribute to the derogation of these books. In addition, a text-based analysis of romance and other book genres could reveal the factors that make them of potentially lower quality (e.g., sentence structure, simplicity of writing, themes, etc.).

Conclusion

Overall, the results of our two studies suggest that there is little evidence of bias against female authors of fiction when readers evaluate textual excerpts, based solely on the attribution of author gender and controlling for the actual text. Further, participants in this research were not biased against romance novels strictly as a function of their association with female authors.

These results represent an important first step in beginning to understand the negative evaluations of female-centered media, such as romance novels. In addition, these results may begin to challenge long-standing beliefs held by publishers about gender biases in readers.

Footnotes

[1] Comparisons between the online and in-person participants, for both studies, can be found in the supplemental materials: https://osf.io/vj2pb/?view_only=0343f5f313914933862cc2ab5f15567f.

[2] While designing this study, a seventh item regarding perceived expertise of the author was discussed, but we ultimately chose not to include this item as it is not so germane to narrative fiction. However, we forgot to remove this item from our preregistration, even though it was not in our final design.

[3] The author names were to be presented again with these questions, however, due to a computer error, this did not happen for participants who completed the study in person.

References

Akbar, A. (2012, May 14). How e-readers took the embarrassment out of erotic fiction. *Independent*. Retrieved from https://www.independent.co.uk/arts-entertainment/books/features/how-e-readers-took-the-embarrassment-out-of-erotic-fiction-7743289.html

Armitage, H. (2018). 12 women writers who wrote under male pseudonyms. Retrieved from https://theculturetrip.com/north-america/usa/articles/12-female-writers-who-wrote-under-male-pseudonyms/

Bauer, D. J. & Serba, S. K. (2001). Fitting multilevel models with ordinal outcomes: Performance of alternative specifications and methods of estimation. *Psychol Methods, 16*(4), 373-390. doi:10.1037/a0025813.

Bookcareers.com salary survey results 2017. (2017). Retrieved from https://www.bookcareers.com/salary-survey-2/bookcareers-com-salary-survey-results-2017/

Bookish. (2014). Fox in the hen house? Romance authors you didn't know are men. Retrieved from https://www.bookish.com/articles/fox-in-the-hen-house-romance-authors-you-didnt-know-are-men/

Brackett, K. P. (2000). Facework strategies among romance fiction readers. *The Social Science Journal, 37*(3), 347–360. https://doi.org/10.1016/S0362-3319(00)00073-2

Christensen, R. H. B. (2019). "ordinal: Regression Models for Ordinal Data." R package version 2019.12-10, URL https://CRAN.R-project.org/package=ordinal.

Ciechanowicz, A. (1983). Effect of author's and subject's gender on perception of author and text. *Polish Psychological Bulletin, 14*(2), 107–114.

Colley, A., North, A., & Hargreaves, D. J. (2003). Gender bias in the evaluation of New Age

music. *Scandinavian Journal of Psychology*, *44*(2), 125–131. DOI: 10.1111/1467-9450.00330

Denham, J. (2015, August 6). Writing under a male name makes you eight times more likely to get published, one female author finds. *Independent*. Retrieved from https://www.independent.co.uk/arts-entertainment/books/news/writing-under-a-male-name-makes-you-eight-times-more-likely-to-get-published-one-female-author-finds-10443351.html

Dienes, Z. (2014). Using Bayes to get the most out of non-significant results. *Frontiers in Psychology*, *5*(781), 1-17. doi:10.3389/fpsyg.2014.00781

Flood, A. (2015). Sexism in publishing: "My novel wasn't the problem, it was me, Catherine." Retrieved from https://www.theguardian.com/books/2015/aug/06/catherine-nichols-female-author-male-pseudonym

Gallivan, J. (1991). Gender bias in students' ratings of essays. *Journal of Social Behaviour and Personality*, *6*(1), 119–124.

Goldberg, P. (1968). Are women prejudiced against women? *Trans-Action*, *5*(28), 28–30. https://doi.org/https://doi.org/10.1007/BF03180445

Haemmerlie, F. M., & Montgomery, R. L. (1991). Goldberg revisited: Pro-female evaluation bias and changed attitudes toward women by engineering students. *Journal of Social Behaviour and Personality*, *6*(2), 179–194.

Kang, S. K., DeCelles, K. A., Tilcsik, A., & Jun, S. (2016). Whitened resumes: Race and self-presentation in the labor market. *Administrative Science Quarterly*, *61*(3), 469–502. https://doi.org/10.1177/0001839216639577

Kaplan, R. M. (1978). Is beauty talent? Sex interaction in the attractiveness halo effect. *Sex*

Roles, *4*(2), 195–204. https://doi.org/10.1007/BF00287500

Koolen, C. W. (2018). Reading beyond the female : The relationship between perception of author gender and literary quality [Doctoral dissertation, University of Amsterdam]. UvA Digital Academic Repository.

Koolen, C., van Dalen-Oskam, K., van Cranenburgh, A., & Nagelhout, E. (2020). Literary quality in the eye of the Dutch reader: The National Reader Survey. *Poetics, 79.* https://doi.org/10.1016/j.poetic.2020.101439

Kuznetsova, A., Brockhoff, P.B., & Christensen, R.H.B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software, 82*(13), 1–26. doi: 10.18637/jss.v082.i13.

Lebuda, I., & Karwowski, M. (2013). Tell me your name and I'll tell you how creative your work is: Author's name and gender as factors influencing assessment of products' creativity in four different domains. *Creativity Research Journal*, *25*(1), 137–142. https://doi.org/10.1080/10400419.2013.752297

Levenson, H., Burford, B., Bonno, B., & Daiv, L. (1975). Are women still prejudiced against women? A replication and extension of Goldberg's study. *The Journal of Psychology*, *89*, 67–71. https://doi.org/10.1080/00223980.1975.9923907

Lois, J., & Gregson, J. (2015). Sneers and leers: Romance writers and gendered sexual stigma. *Gender & Society*, *29*(4), 459–483. https://doi.org/10.1177/0891243215584603

Milliot, J. (2019, November 15). The PW publishing industry Salary Survey. *Publishers Weekly*. Retrieved from https://www.publishersweekly.com/pw/by-topic/industry-news/publisher-news/article/81718-the-pw-publishing-industry-salary-survey-2019.html

Mischel, H. (1974). Sex bias in the evaluation of professional achievements. *Journal of*

Educational Psychology, *66*(2), 157–166. https://doi.org/10.1037/h0036283

Moore, M., & Trahan, R. (1998). Evaluating an excerpt about gender: Does sex of an author matter? *Psychological Reports*, *82*, 247–253. https://doi.org/10.2466/pr0.1998.82.1.247

Nobel Media. (2019). All Nobel Prizes in literature. Retrieved from https://www.nobelprize.org/prizes/lists/all-nobel-prizes-in-literature/

Owen, M. (1997). Re-inventing romance: Reading popular romantic fiction. *Women's Studies International Forum*, *20*(4), 537–546. https://doi.org/10.1016/S0277-5395(97)00042-3

Owen, P. (2018). Women-fronted films are top three highest-grossing movies of 2017. Retrieved from https://www.theguardian.com/film/2018/jan/01/women-fronted-films-top-three-highest-grossing-2017

Paludi, M. A., & Bauer, W. D. (1983). Goldberg revisited: What's in an author's name. *Sex Roles*, *9*(3), 387–390. https://doi.org/10.1007/BF00289673

Peck, T. (1978). When women evaluate women, nothing succeeds like success: The differential effects of status upon evaluations of male and female professional ability. *Sex Roles*, *4*(2), 205–213. https://doi.org/10.1007/BF00287501

Pirri, C., Eaton, E., & Durkin, K. (1995). Australian professional women's evaluations of male and female written products. *Sex Roles*, *32*(9), 691–697. https://doi.org/10.1007/BF01544219

Prasad, R. (2018). Is a female lead now key to box office success? Retrieved from https://www.bbc.com/news/world-us-canada-46543086

Morey, R. D., Rouder, J. N., Jamil, T., Urbanek, S., Forner, K., & Ly, A. (2018). BayesFactor: Computation of Bayes Factors for Common Designs. R package version 0.9.12-4.2. https://CRAN.R-project.org/package=BayesFactor.

Romance Writers of America. (n.d.). About the romance genre. Retrieved October 31, 2018, from https://www.rwa.org/p/cm/ld/fid=578

Schneider, M. (2019). 100 most-watched TV shows of 2018-19: Winners and losers. Retrieved from https://variety.com/2019/tv/news/most-watched-tv-shows-highest-rated-2018-2019-season-game-of-thrones-1203222287/

Shevenock, S. (n.d.). Reality is America's least favourite TV genre - Yet people are still watching. Retrieved from https://morningconsult.com/2018/11/27/reality-is-americas-least-favorite-tv-genre-yet-people-are-still-watching/

Statisitca. (2015). What types of books have you read in the past year? Retrieved from https://www.statista.com/statistics/201404/types-of-books-that-american-adults-read/

Swim, J., Borgida, E., Maruyama, G., & Myers, D. G. (1989). Joan McKay versus John McKay: Do gender stereotypes bias evaluations? *Psychological Bulletin, 105*(3), 409–429. https://doi.org/10.1037/0033-2909.105.3.409

Watson, A. (2019). Genre breakdown of the top 250 TV programs in the United States in 2017. Retrieved from https://www.statista.com/statistics/201565/most-popular-genres-in-us-primetime-tv/

Weinberg, D. B., & Kapelner, A. (2018). Comparing gender discrimination and inequality in indie and traditional publishing. *PloS one, 13*(4), 1-20. https://doi.org/10.1371/journal.pone.0195298

Zhang, J., Hennessy, D. A., Luo, J., Song, Y., Ren, K., Zhang, Q., Han, Z., & Yao, P. (2009). Are women in china sexist toward other women? A study of Chinese college students. *Psychological Record, 105*(1), 267–274. https://doi.org/10.2466/PR0.105.1.267-274

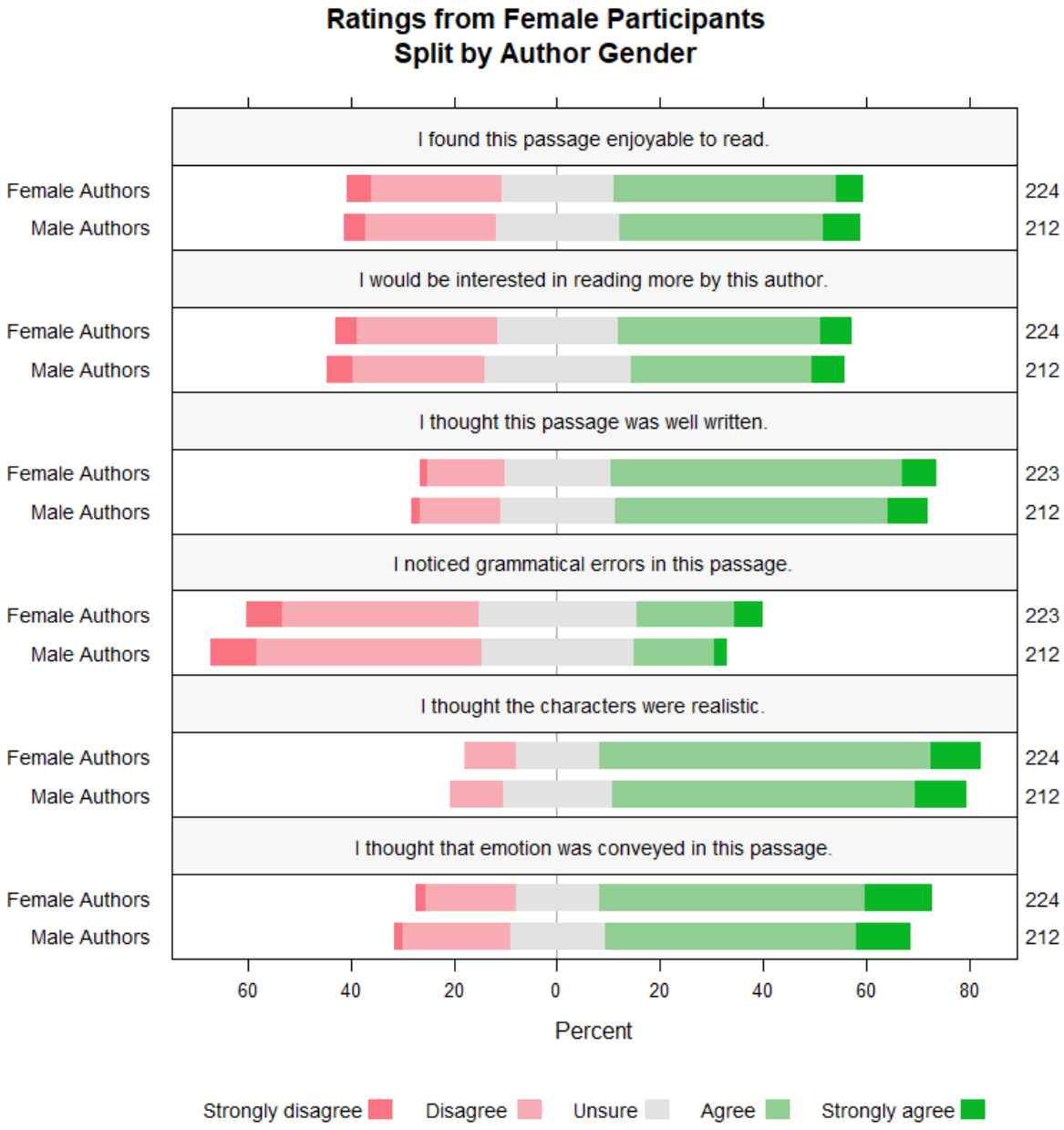
Table 1.
Study 1 Bayes Factors for each factor in all analyses

	Dependent Measures						
Fixed Effects	Liking	Errors	Well-written	Realistic	Emotion	Read more	Total
Participant Gender	.31	.41	.17	.15	.13	.93	.16
Author Gender	.25	.49	.09	.09	.09	.11	.17
A. Gender X P. Gender	.03	.04	.002	.01	.003	.06	.01

Table 2.
Study 2 Bayes Factors

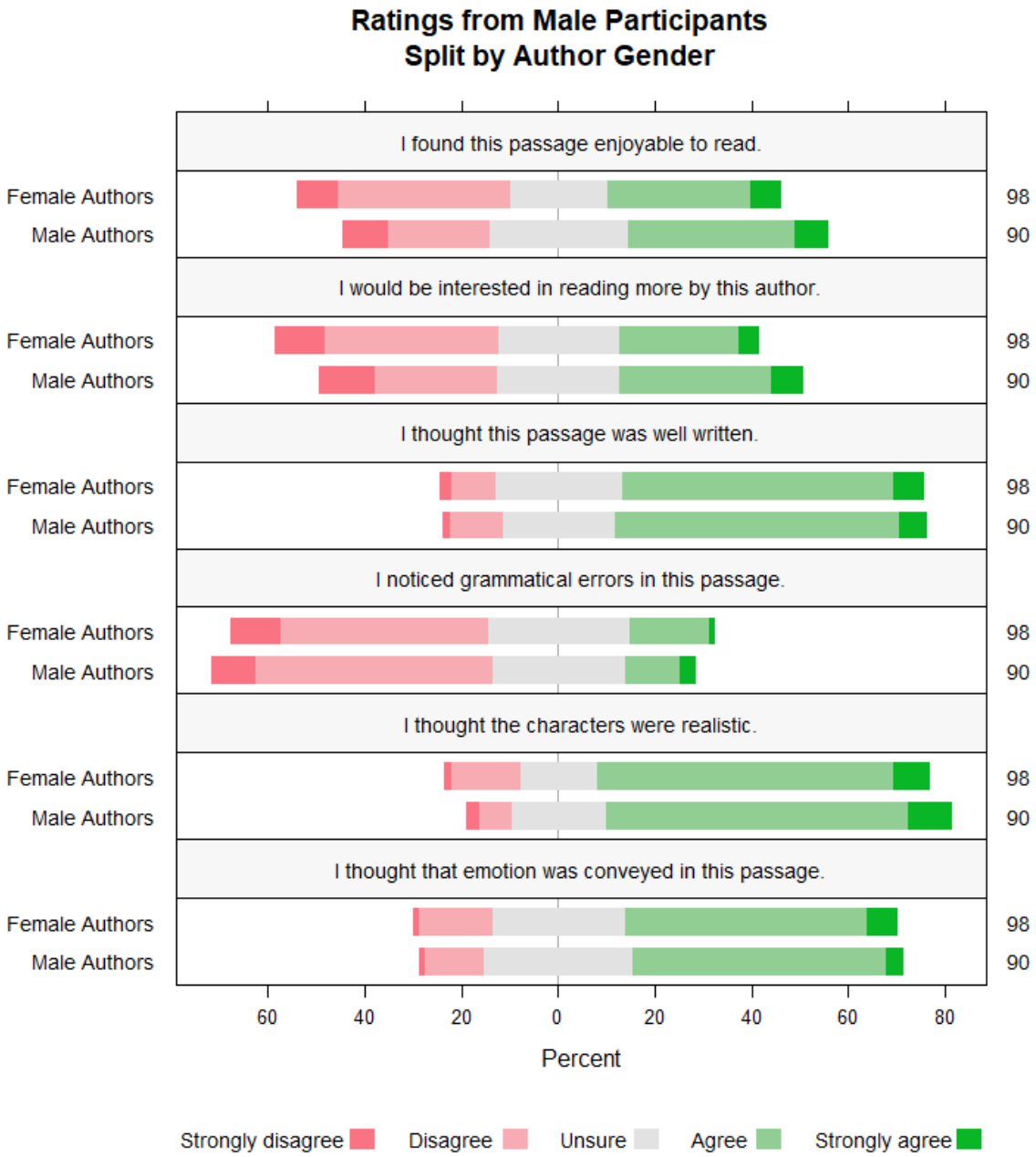
	Dependent Measures						
Fixed Effects	Liking	Errors	Well-written	Realistic	Emotion	Read more	Total
Participant Gender	.23	.88	.16	.20	.15	.69	.15
Author Gender	.35	.11	.11	.26	.15	.14	.23
Book Genre	.30	.16	.14	.17	.55	.15	.21
A. Gender X P. Gender X Genre	<.001	<.001	<.001	<.001	<.001	<.001	<.001

Figure 1.
Study 1 Effect of Author Gender for Female Participants



Note. The percentage of female participants who responded with each of the 5 response options (color-coded, as indicated in the legend) appears along the x-axis. These responses are broken down by the author gender attributed to the passage along the y-axis, along with the number of participants for each group.

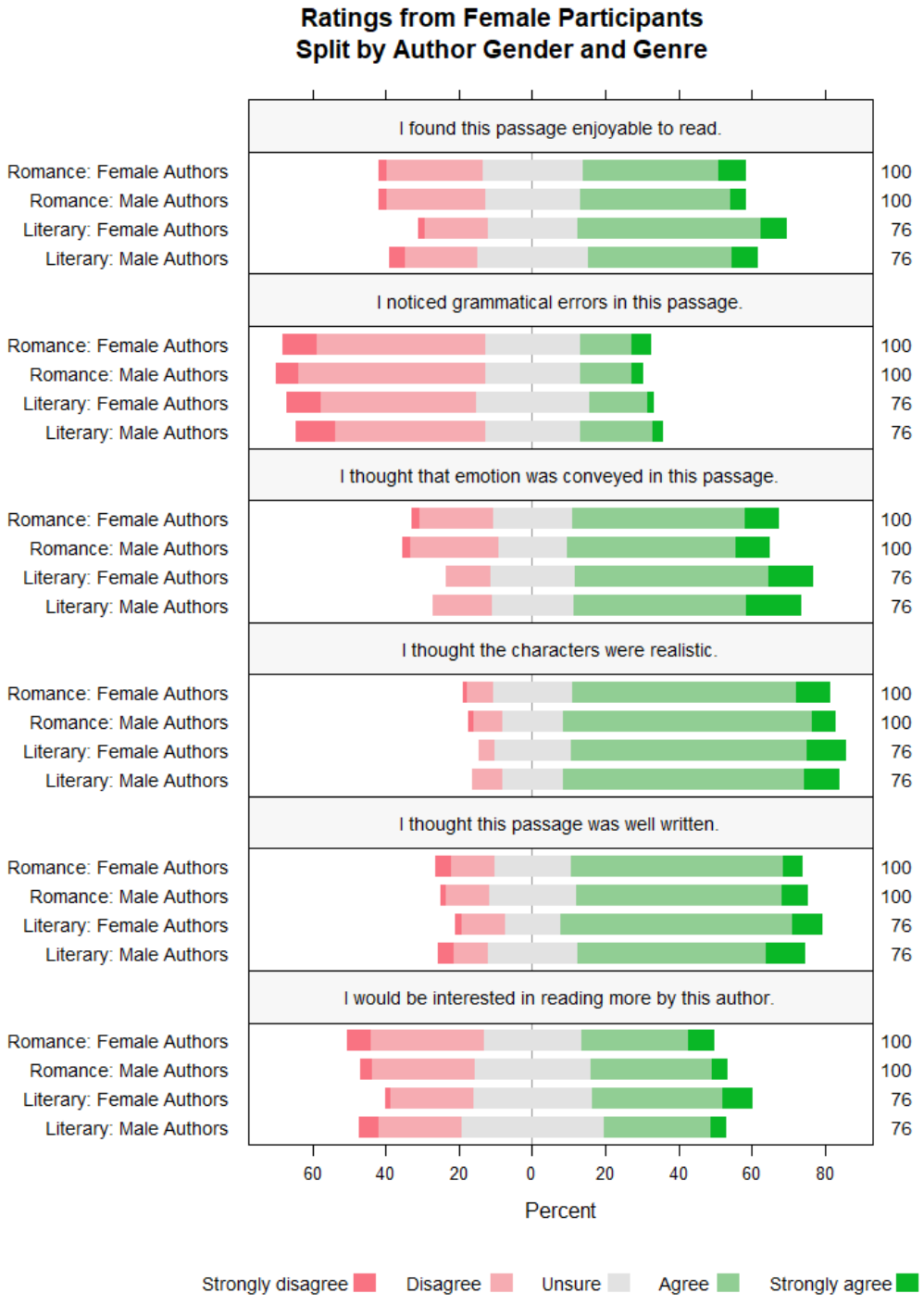
Figure 2.
Study 1 Effect of Author Gender for Male Participants



Note. The percentage of male participants who responded with each of the 5 response options (color-coded, as indicated in the legend) appears along the x-axis. These responses are broken down by the author gender attributed to the passage along the y-axis, along with the number of participants for each group.

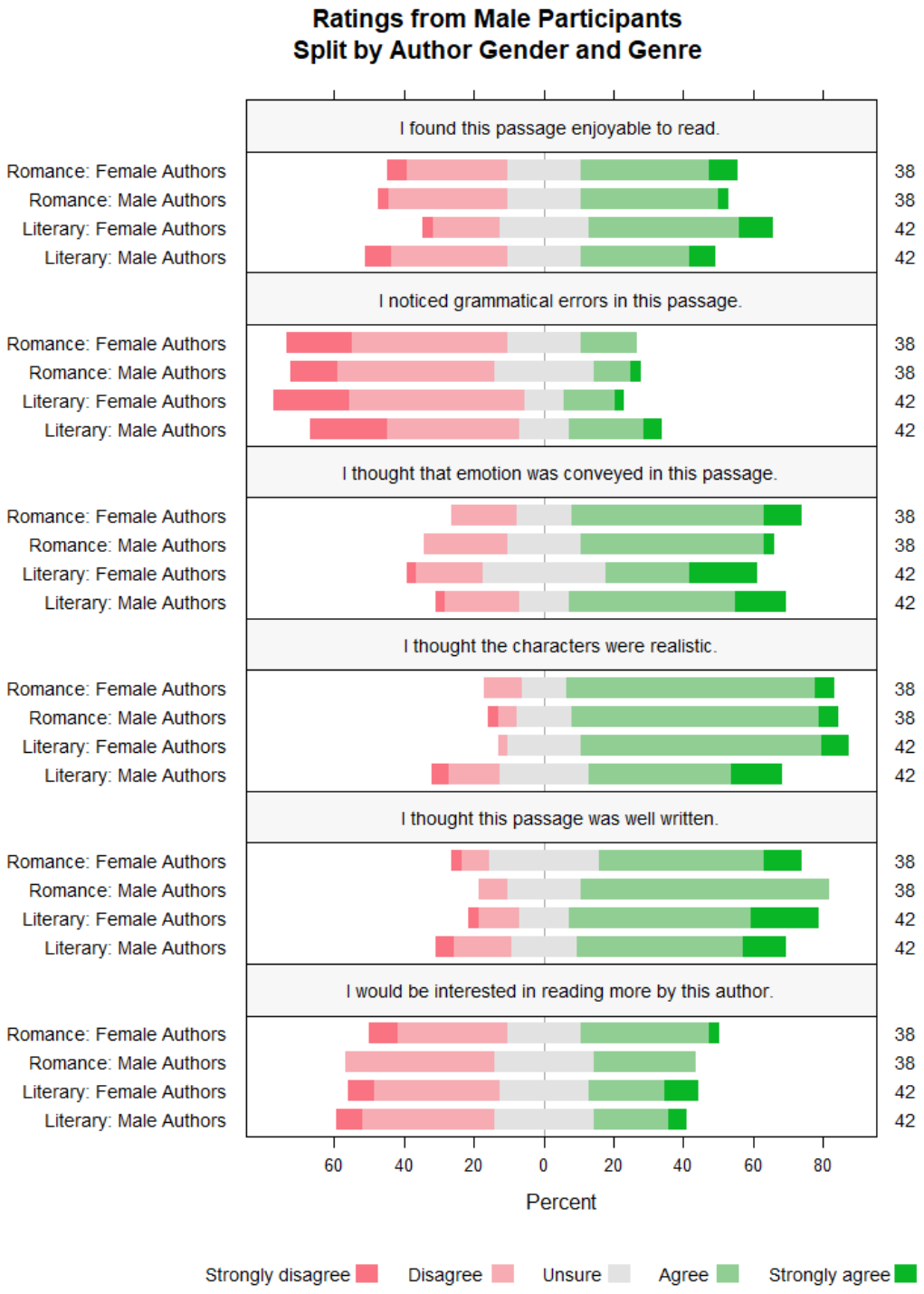
Figure 3.

Study 2 Effect of Author Gender and Book Genre for Female Participants



Note. The percentage of female participants who responded with each of the 5 response options (color-coded, as indicated in the legend) appears along the x-axis. These responses are broken down by the genre and author gender attributed to the passage along the y-axis, along with the number of participants for each group.

Figure 4.
Study 2 Effect of Author gender and Book Genre for Male Participants



Note. The percentage of male participants who responded with each of the 5 response options (color-coded, as indicated in the legend) appears along the x-axis. These responses are broken down by the genre and author gender attributed to the passage along the y-axis, along with the number of participants for each group.