2018

Gender Bias in Medical Images Affects Students' Implicit but not Explicit Gender Attitudes

Rhiannon Parker
University of New South Wales, rbp774@uowmail.edu.au

Theresa A. Larkin
University of Wollongong, tlarkin@uow.edu.au

Jonathan P. Cockburn
University of Wollongong, jon@uow.edu.au

Publication Details
Gender Bias in Medical Images Affects Students' Implicit but not Explicit Gender Attitudes

Abstract
Medical education curricula have the potential to impact the gender attitudes of future healthcare providers. This study investigated whether gender-biased imagery from anatomy textbooks had an effect on the implicit and explicit gender attitudes of students. We used an online experimental design in which students (N = 456; 55% female) studying anatomy were randomly assigned to a visual priming task using either gender-neutral or gender-biased images. The impact of this priming task on implicit attitudes was assessed using the Implicit Association Test (IAT) and the impact on explicit attitudes was measured using the Gender Bias in Medical Education Scale. Viewing biased images was significantly positively associated with implicit gender bias as indicated by higher IAT scores in the treatment compared to the control condition (mean IAT difference = 43 milliseconds; Cohen’s d = .33). In contrast, there was no significant effect of gender-biased images on explicit gender attitudes.

Keywords
implicit, but, gender, not, bias, explicit, attitudes, medical, images, affects, students’

Disciplines
Arts and Humanities | Law

Publication Details

This journal article is available at Research Online: https://ro.uow.edu.au/lhapers/3656
Introduction

Gender bias is prevalent in the medical domain and is associated with healthcare disparities (Chiaramonte et al., 2008; Hamberg, 2008; Verdonk, Benschop, De Haes, & Lagro-Janssen, 2009). For example, the exclusion of females from medical trials likely contributes to women with heart disease remaining underdiagnosed, undermanaged, and undertreated (Mehta et al., 2016). The biases of healthcare practitioners, in particular, play a significant role in contributing to these disparities (Chapman, Kaatz, & Carnes, 2013; White, 2011). The persistent use of gender stereotypes by physicians, for example, has meant that women are more likely than men to have their reports of pain be perceived as emotional (Hoffmann & Tarzian, 2001), and to be incorrectly diagnosed with psychological rather than somatic disorders (Smith, 2011). One of the primary arenas where healthcare providers’ gender attitudes can be influenced is during medical education, where curriculum and teaching practices can contribute to the development and maintenance of gender ideologies (Bickel, 2001; Stromquist, Lee, & Brock-Utne, 2013). As such, strategies aimed at preventing gender bias in healthcare need to start at the educational level (Hamberg, 2008; Risberg, Johansson, & Hamberg, 2009; Teal et al., 2010; Verdonk et al., 2009). There is therefore a need to examine the role that a medical education curriculum plays in the formation of gender-biased attitudes.

Empirical research has highlighted the importance of attitudes by identifying the way in which they manifest themselves through behavior (Maio & Haddock, 2009; Smith, Mackie, & Claypool, 2015; Vaughan & Hogg, 2008). Studies have shown that the attitudes of healthcare practitioners can have an impact on medical treatment and care (Blair et al., 2013; Phelan et al., 2015; Stepanikova, 2012). For example, physicians who exhibit gender-biased attitudes are less likely to recommend women with the same symptoms as men for cardiovascular testing (Daugherty et al., 2017). Attitudes are understood as existing at both an explicit (conscious) and implicit (unconscious) level (Gawronski & Payne, 2010), and individuals can hold implicit attitudes that conflict with their consciously approved beliefs (Greenwald & Banaji, 1995; Wilson, Lindsey, & Schooler, 2000). It is well established that explicit attitudes can have an influence on behavior (Armitage & Conner, 2001). However, not only do implicit attitudes exert a powerful influence on behavior, they are also more likely to predict discriminatory behavior than explicit reports (Gawronski & Payne, 2010; O’Brien et al.,
Effects of Images on Gender Attitudes

Participants were students from two different strands of study within the School of Medicine at the University of Wollongong, Australia: undergraduate medical and health science students enrolled in a first-year anatomy subject (SHS111) and graduate-entry medical students in the first year of their integrated degree (MEDI601). There were 491 students enrolled in SHS111 and 82 students enrolled in MEDI601, for 573 possible participants.

Study Design and Implementation

The study was conducted during anatomy laboratory classes over two weeks (April–May 2015), after ethics approval was obtained from the University of Wollongong Human Research Ethics Committee (ethics approval number HE14/130). Participation was voluntary, and all data collected was anonymous. To reduce distractions and provide students with a convenient way to complete the study, we set up the online task in a private computer room adjacent to the anatomy laboratory. Groups of up to eight students at a time took part and the task took approximately 10 minutes to complete. A researcher remained in the room during all data collection times so that questions or concerns could be directly addressed.

The study used a randomized control trial (RCT) in the form of an online task that was developed for the current study. The online task consisted of four parts: (a) a short introduction that described and provided instructions for the study; (b) a priming task during which participants were randomly assigned by a random number generator embedded in the task to view either gender-biased images for the treatment group or gender-neutral images for the control group; (c) the Implicit Association Test (IAT); and (d) the Gender Bias in Medical Education Scale (GBMES). As participants were unaware of the condition they were assigned to, and as we used a fully automated assignment and treatment procedure, our research used a double-blind random assignment procedure.

Materials

Priming Task. For the priming task, gender-biased images for the treatment condition and gender-neutral images for the control condition were selected from existing anatomy textbooks. Specifically, the treatment condition included images of males in traditionally masculine sports roles (e.g., men playing soccer) and females in traditionally feminine reproductive roles (e.g., a mother breastfeeding). The control condition consisted of images of internal anatomy in which the sex or gender could not be determined, including isolated organs such as the lung, heart, and brain, and musculoskeletal structures such as the vertebral column and muscles. Twenty-four images were displayed on the screen in automatic succession for four seconds each, running for 96 seconds. To increase participants’ engagement with the priming task, the question “What anatomy chapter would this image appear in?” appeared below each image.

Implicit Measures. Following the visual priming task, participants completed an IAT, adapted from https://github.com/winteram/IAT (Mason, Allon, & Ozturk, 2018). The
IAT measured the categorization speed at which participants associated stereotypically similar (i.e., reproductive health terms with female pronouns and sports health terms with male pronouns) and dissimilar (i.e., reproductive health terms with male pronouns and sports health terms with female pronouns) terms (see online supplemental material 1 for more details). The categories of reproductive and sports health were chosen as these represent stereotypically female and male roles respectively (Collins, 2011; Haines, Deaux, & Lofaro, 2016; Hardin & Greer, 2009). Terms were vetted for relevance by the second author (an anatomy lecturer) to ensure that they were of approximately equal relevance to men and women despite being stereotypically related to one gender. Thus, while issues like fertility/infertility may be stereotypically related to women they are relevant to men also. Indeed, infertility rates in the population are similar among men and women (see Chandra, Copen, & Stephen, 2013).

After instructions, participants completed the IAT. We used a standard IAT design (Mason et al., 2018). A screenshot of each round and an explanation of what participants were required to do is presented in online supplemental material 2. Test terms were presented in the center of the computer screen. Categories were presented on the top left and top right of the screen (see Figure 1). When the test term in the center of the screen matched a category on the left participants were asked to press the “I” button on their keyboard as quickly as possible. When the test term matched a category on the right participants were asked to press the “E” button as quickly as possible. The standard procedure consisted of seven rounds broken into two main sections, each with familiarization, practice, and recorded rounds. The first section (rounds one to four) aimed to gain a baseline estimate of participants’ speed in co-categorizing stereotypically similar gender and medical terms (i.e., female pronouns and reproductive health terms were presented as categories on the left; male pronouns and sports health terms were presented on the right). Round one asked participants to assign gender terms only (female on the left; male on the right). Round two asked participants to assign health terms only (reproductive health on the left; male on the right). Round three, presented in Figure 1a, was a practice round with participants assigning intermixed gender and health terms to categories on the left (male pronouns and reproductive health terms) or the right (female pronouns and sports health terms). Round six was identical to round six, but here participants’ average response times were recorded in milliseconds. The IAT score, which was used as the primary outcome in this research, was taken by subtracting participants’ average response times in round seven from those in round four. The participant average response time difference in milliseconds and the participant average response time difference in milliseconds corrected by the standard deviation were measured. The latter resulted in a slightly bigger effect size for the treatment group. However, the average categorizing speed in milliseconds was retained as the primary scaling of the dependent variable to provide a more straightforward interpretation of results. Responses less than 300 or greater than 3000 milliseconds long were excluded as invalid responses.

Explicit Measures. Explicit attitudes were measured using the GBMES, a three-factor measure of participants’ explicit attitudes toward gender bias in medical education (Parker, Parker, Larkin, & Cockburn, 2016). The GBMES consists of 10 items using a six-point Likert scale, ranging from 0 (strongly disagree) to 5 (strongly agree) (see online supplemental material 3). It measures three factors related to gender bias: (a) awareness, for example, “In anatomy textbooks, reproductive chapters have more images of females than males”; (b) beliefs, for example, “I believe educators should raise awareness of the risks of gender bias in medicine”; and (c) experiences, for example, “I have seen evidence of gender bias in anatomy class activities.” We used the scale scores (i.e., the average item score for a given factor) in this analysis. Thus, average agreement for a GBMES factor was represented by a score of 2.5. The coefficient of reliability for the three scale scores of the GBMES ranged from .72 to .91.

Analysis

Group differences between the control and treatment conditions on the implicit and three explicit attitude dependent variables were assessed using t-tests in R (R Core Development Team, 2014). Results are given in the original metric and in the form of Cohen’s d, indicating differences between groups in standard deviation units. A sensitivity analysis using ANCOVA was undertaken, in which we tested the difference between the treatment condition and the control condition on the outcomes, controlling for key covariates (i.e., age, gender, and program of study). This is referred to as a doubly robust design in which both random assignment and statistical adjustment are used to mitigate potential biasing
selection effects and thus construct robust counterfactuals for causal inference (Morgan & Winship, 2015).

**Results**

**Demographics**

Out of the 573 students invited to participate, 456 voluntarily participated in this study (response rate 80%). This included 252 females (55%) and 190 males (42%), with a mean age of 20.7 ± 4.5 years (median = 19 years). From the undergraduate cohort, 384 out of 491 participants took part (76% female), with a mean age of 19.9 ± 4.1 years (median = 18 years), and 91% were in their first year of university. Among the graduate-entry medical students, 72 out of 82 participants took part (37% female), with a mean age of 24.9 ± 4.1 years (median = 23 years). Random assignment of participants to the image groups for the priming task resulted in 248 in the treatment (56% female) and 208 in the control (54% female) condition.

In the methods section we assumed that participants would find it harder to assign stereotypically dissimilar terms (e.g., male pronouns and reproductive terms) to a common target than stereotypically similar terms (e.g., male pronouns and sports injury terms). This was indeed the case, with participants taking 168 milliseconds longer to assign terms in the stereotypically dissimilar condition (paired t-test: $t_{[455]} = 27.149, p < .001$).

**Randomized Control Trial Results**

Results indicated that the difference in categorization speed for the IAT between the treatment and control conditions was significant (Table 1), with those in the treatment condition being on average 36 milliseconds slower in the categorization tasks for stereotypically dissimilar than similar groups compared to those in the control condition. This effect size was small ($d = .28$) according to standard metrics in psychology (Cohen, 1992). Controlling for key covariates—gender, age, and program of study—via ANCOVA altered the effect size little, with those in the treatment condition being 43 milliseconds slower than those in the control condition ($d = .33$). For the explicit measures from the GBMES there was no significant difference in attitudes between those in the treatment condition being 43 milliseconds slower than those in the control condition ($d = .33$). For the explicit measures from the GBMES there was no significant difference in attitudes between those in the treatment condition being 43 milliseconds slower than those in the control condition ($d = .33$). For the explicit measures from the GBMES there was no significant difference in attitudes between those in the treatment condition being 43 milliseconds slower than those in the control condition ($d = .33$). For the explicit measures from the GBMES there was no significant difference in attitudes between those in the treatment condition being 43 milliseconds slower than those in the control condition ($d = .33$). For the explicit measures from the GBMES there was no significant difference in attitudes between those in the treatment condition being 43 milliseconds slower than those in the control condition ($d = .33$). For the explicit measures from the GBMES there was no significant difference in attitudes between those in the treatment condition being 43 milliseconds slower than those in the control condition ($d = .33$). For the explicit measures from the GBMES there was no significant difference in attitudes between those in the treatment condition being 43 milliseconds slower than those in the control condition ($d = .33$).

**Discussion**

This experimental study investigated whether gender-biased imagery influenced the implicit and explicit gender attitudes of students studying anatomy. The results showed...
that viewing gender-biased images had a significant effect of moderate size on implicit gender attitudes. This is consistent with research showing that implicit attitudes are contextually sensitive and can be influenced by environmental factors (Barden et al., 2004; Wittenbrink, 2007). These results are significant given that implicit bias has been linked with issues such as poorer patient care (Blair et al., 2013; Cooper et al., 2012) and treatment decisions (Stepanikova, 2012). Thus, the visual representation of bias in educational content may have the potential to impact healthcare outcomes.

That exposure to gender-biased images had a statistically significant effect on implicit bias is important given that the treatment was short (96 seconds). Research has shown that medical textbooks, a core part of medical education, are replete with visual gender bias (Giacomini et al., 1986; Lawrence & Bendixen, 1992; Mendelsohn et al., 1994; Parker et al., 2017). Thus, the results of this paper give rise to questions about the implications of long-term exposure to gender bias in visual images in medical curricula. It seems possible that exposure to biased images throughout students’ medical education career may cumulatively influence implicit attitudes. Indeed, some research has suggested that long-term exposure to news stereotypes not only has a significant impact on implicit attitudes but that this, in turn, modifies explicit attitudes (Arendt & Northup, 2015). Future research could explore the duration of the effects of a priming task and the effects of long-term exposure to gender-biased imagery in medicine.

In contrast to the significant effect of viewing gender-biased images on implicit attitudes, there was no observed effect of the treatment condition on participants’ explicit reports of bias in the GBMES. This is noteworthy as research has shown that implicit attitudes are more likely to predict disparities in medical treatment and care than explicit attitudes (Green et al., 2007; Stepanikova, 2012). Such a result is in line with existing theories that individuals can simultaneously hold disparate implicit and explicit attitudes (Wilson et al., 2000). However, this disparity could also be explained by that fact that explicit attitudes can be much easier to disguise than implicit attitudes ((Blair, Steiner, & Havranek, 2011; Dovidio, Kawakami, & Beach, 2001).

**Limitations**

Readers should be aware of some limitations to the generalization of the findings. The participants were all from a single university and their attitudes and experiences may differ from other subsets of the population. Although this study provides evidence of the influence of gender-biased images on implicit attitudes, it remains unclear how long these effects last. This study also focused solely on the effects of images as a source of gender bias; it would be beneficial to explore how other curricula’s content and educational experiences may contribute to or counteract gender-biased attitudes. We used a gold standard double-blind random assignment procedure to assign participants to conditions. With 491 participants, we also had adequate power to not only detect a significant effect but also to ensure balance across key covariates. Indeed, the sensitivity analysis in which we conditioned on key covariates resulted in little change to the estimated treatment effect. While these design features and sensitivity results suggested we did indeed have an unbiased treatment effect, we cannot guarantee that there were not pre-existing differences between the treatment and control group in underlying bias prior to assignment. While there are concerns about practice effects, future research should consider collecting pre-treatment data on implicit and explicit bias. Despite these limitations, the results of this study have significant implications for the presence of visual gender bias within medical curricula.

**Conclusion**

The current study demonstrated that viewing gender-biased images can have a significant impact on the implicit gender attitudes of students studying anatomy. This shows that implicit attitudes are affected by context and can be influenced via interventions such as visual priming tasks. A
medical curriculum design that avoids both visual and textual gender bias and which aims to provide bias-reduction interventions may ultimately improve healthcare outcomes. Studies such as this are therefore important for providing medical educators with information on why and how to reduce implicit gender bias during medical education.

Availability of Data
The dataset supporting the conclusions of this article is available in the figshare repository: https://figshare.com/articles/Study3_Data/3798228

Funding
This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

References


Hardin, M., & Greer, J. D. (2009). The influence of gender-role socialization, media use and sports participation on perceptions


### Authors

RHIANNON B. PARKER is a postdoctoral research fellow at the Practical Justice Initiative, Centre for Social Research in Health, University of New South Wales. She specializes in the sociology of health and illness, exploring the social dynamics of cancer and palliative care and investigating the role that medical education plays in constructing knowledge about gender and the body.

THERESA LARKIN is a senior lecturer in medical sciences at the Graduate School of Medicine, University of Wollongong. Her research focuses on the roles of oxytocin and cortisol in health and disease, gender differences in anatomy and physiology, and applying the cognitive theory of learning in medical and anatomical education.

JON COCKBURN is a design theory lecturer, research degree supervisor, and course coordinator in the Creative Industries Group in the Faculty of Law, Humanities and the Arts at the University of Wollongong, Australia. His research focuses on fashion, media, and visual communications theory using humanities models of scholarship.