Investigating User Perception of Gender Bias in Image Search: The Role of Sexism

Jahna Otterbacher*, Alessandro Checco**, Gianluca Demartini***, Paul Clough**

Open University of Cyprus, Nicosia CYPRUS*, University of Sheffield, Sheffield, UK**, University of Queensland, St. Lucia, AUSTRALIA***

Introduction

One means to promote greater transparency of search algorithms – which are typically complex and proprietary – is to raise user awareness of biased result sets. Currently, we present the results of a *controlled experiment via crowdsourcing* using participants recruited from three countries to measure the extent to which workers perceive a given results set to be subjective or objective. Demographic information about the workers, along with measures of sexism, were collected. Amongst other findings, the results confirm that **sexist people are less likely to detect and report gender biases** in image search results.

According to social psychologists, sexism is characterized by an ambivalence towards women: people may be hostile towards women or they may hold benevolent attitudes, stereotyping them into limited, traditional roles [3]. The Ambivalent Sexism Inventory (ASI), consisting of 22 Likert items, has been shown to reliably tap the two components of sexism: Hostile (HS) and Benevolent (BS) [3]. While the HS and BS scores are positively correlated, HS tends to be associated with holding negative stereotypes of women, and BS tends to predict positive – although traditional – views of women. In addition, [4] demonstrated that HS and BS are universal across cultures. However, sexism is related to culture, and national averages on these measures are correlated to the levels of gender inequality in the society.

Research Questions

We use the ASI as a tool for understanding how users perceive gender bias in image search engine results. We presented participants with images returned by Google in response to character trait searches. As shown in Figure 1, extrapolating from theory, we expect that a user's level of sexism will correlate to the manner in which she perceives results, which in turn, influences the evaluation of the results. Specifically, we address two research questions:

(1) Are sexist/non-sexist people less/more likely to evaluate a heavily gender-imbalanced result set as being subjective?

(2) Is there evidence that sexist/non-sexist people perceive a given image result set differently?

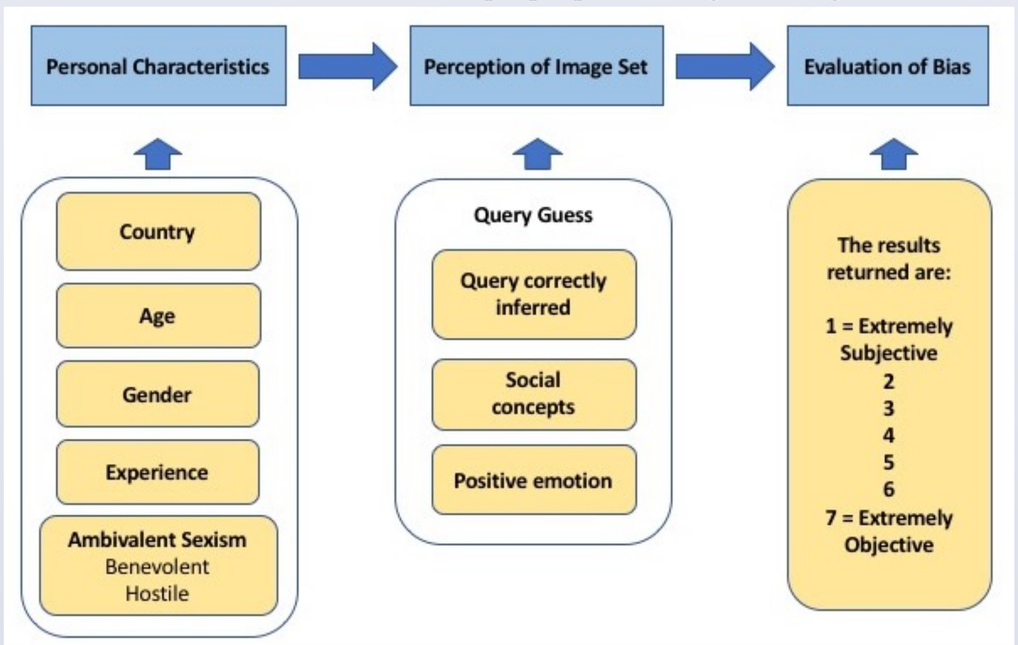


Figure 1: Conceptual model.

Query	smart person	aggressive person	warm person	anxious person	hot air balloon
Trait	+	-	+	-	=
Bias	M	M	F	F	na

Table 1: Query used with corresponding trait (+ for positive, - for negative) and bias (M/F for bias towards males/females).

Query	# Non-objective	Country	Age	Gender	log(Experience)	Benevolent	Hostile	Pseudo R^2
Hot air balloon	28	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	0.05
Smart person	70	n.s.	n.s.	n.s.	n.s.	-0.773**	n.s.	0.17
Aggressive person	59	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	0.06
Warm person	92	n.s.	n.s.	-0.690*	-0.247*	-0.510*	n.s.	0.13
Anxious person	60	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	0.04

Table 2: Logistic regression models to predict event where image result set is evaluated as "not objective."

Methodology

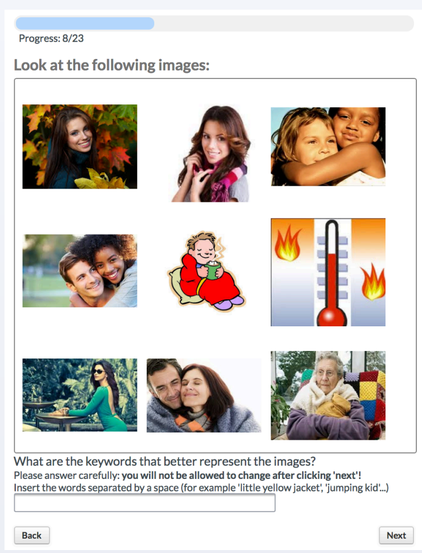
We conducted an experiment on the Crowdflower platform, repeating the following procedure with crowdworkers located in located in the UK, US, and India. In total, 281 users participated.

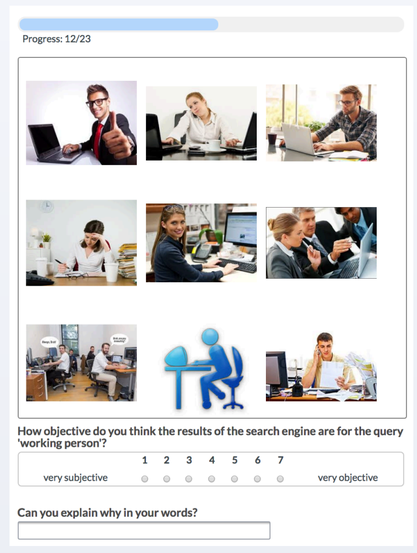
Part 1 (guess the query): We show users a grid of 9 images and ask them what keywords best describe them. The question is repeated for multiple grids, for the queries shown in Table 1. The phrasing of the question does not disclose the fact that the image grid has been generated via an image search engine.

Part 2 (search engine opinions): Users are then asked to answer questions regarding the objectivity of search engines, along with several proficiency self-assessment questions.

Part 3 (perceived bias): Now, the user is told that the grid of images was obtained from a search engine. For each image grid, the query is disclosed and the user is required to compare it to the answer provided in Part 1, assessing the objectivity of each image result.

Part 4 (ASI): The user completes the ASI questionnaire to assess his/her level of sexism.





Analysis

Is sexism correlated to the evaluation of the results?

Table 2 presents a logistic regression analysis in which we predict the event that a participant evaluates the image set as *not being objective* (i.e., rating of 1, 2, 3 or 4), based only on their demographic characteristics and the two ASI dimensions.

- On the neutral query, few participants indicated that results were subjective and no participant characteristics were correlated to this outcome.
- For the two image sets based on positive characteristics ("warm" and "smart"), even when user attributes are controlled for, *Benevolent Sexism* is negatively correlated to a participant having reported the image results as not objective.
- In other words, Benevolent sexists, who hold traditionally stereotypical views of women, were less likely than others to view the heavily skewed results sets as being subjective, as compared to other participants.

Do sexists perceive results differently than non-sexists?

We tested the conceptual model depicted in Figure 1, by analyzing the query guess provided in Part 1 of the experiment, for the two queries "smart" and "warm person." We processed the query guess using the Linguistic Inquiry and Wordcount tool (LIWC), evaluating each guess as to the extent to which it 1) matched the true query, 2) contained social concepts, and 3) used words with positive emotion. For each query, we generated a Structural Equation Model, as presented in Table 3.

For both queries, we observe a positive correlation between the latent variable comprising the participant characteristics, and the manner in which she perceived the results set. However, we observe a negative correlation between the latent variable comprising the perception measures and the evaluation measure. However, it should be noted that the correlations, while statistically significant, are weak.

Measurement Model				
	Smart		Warm	
	Est.	*z*	Est.	*z*
→User characteristics				
Male	1.00	fixed	1.00	fixed
Benevolent	3.805	4.420**	3.629	4.492**
Hostile	4.878	4.203**	5.219	3.849**
→Perception				
Guess match	1.00	fixed	1.00	fixed
Social	0.850	7.428**	0.815	3.284*
Positive	0.527	5.490**	0.275	2.646*
→Evaluation				
Objectivity rating	1.00	fixed	1.00	fixed
Structural Model				
User→Perception	0.013	2.628**	0.009	2.155*
Perception→Evaluation	-0.025	-2.535*	-0.028	-2.677**

Table 3: Structural Equation Models.

Conclusions

- Understanding prejudices and beliefs is critical in better understanding how people engage with and evaluate search technologies.
- Our findings confirm that people who test as being more sexist on the ASI measure are less likely to recognize gender bias in image search. In addition, the Benevolent dimension of sexism is an important factor, rather than the Hostile dimension.
- This reinforces stereotypes and may increase the retrievability of some images, at the expense of others.

References

[1] M. Kay, C. Matuszek, S.A. Munson. Unequal representation and gender stereotypes in image search results for occupations. In Proceedings of ACM CHI (2015), 3819-3828.

[2] J. Otterbacher, J. Bates and P. Clough. Competent men and warm women: Gender stereotypes and backlash in image search results. In Proceedings of ACM CHI (2017), 6620-6631.

[3] P. Glick and S. T. Fiske. The ambivalent sexism inventory: Differentiating hostile and benevolent sexism. Journal of personality and social psychology 70, 3 (1996), 491.

[4] P. Glick, S. T Fiske, A. Mladinic, et al. Beyond prejudice as simple antipathy: hostile and benevolent sexism across cultures. Journal of personality and social psychology 79, 5 (2000), 763.

Contacts

The dataset is available at github.com/AlessandroChecco/gender_bias

Authors: jahna.otterbacher@ouc.ac.cy, a.checco@sheffield.ac.uk