Conceptual Centrality and Gender Biases: An Empirical Investigation

Guillermo Del Pinal, Alex Madva, and Kevin Reuter

Abstract

Discussions of socially-relevant biases—including explicit and implicit gender biases—have overlooked an important way in which biases can be encoded in conceptual structures. Most accounts of implicit bias focus on 'associations' between features and the concepts which we use to represent certain categories. Associative properties include things such as cue validity, typicality, and saliency. However, philosophers and cognitive scientists have argued that concepts also encode dependency relations between features and dimensions. Dependency relations determine the conceptual centrality of a feature f, relative to a concept C: i.e., the degree to which other features of C depend on f. Crucially, centrality and measures like cue validity, typicality and saliency can disassociate. In this paper, we argue that dependency between features is one way in which significant biases are encoded in conceptual structure. We will defend this claim by presenting a series of experiments which explore the way in which a particularly important gender bias is encoded.

Keywords: gender bias; conceptual centrality; prototypes; sterotypes; implicit bias
Word count:

1 Introduction

In a very influential recent discussion, Leslie et al. (2015) present a wealth of evidence to argue that women are underrepresented in academic fields whose members believe that brilliance is the main requirement for success. They argue that there are many potential mechanisms by which such field-specific brilliance beliefs may influence women's representation. One of this mechanisms depends on the there being a stereotype such that women (and certain minorities) are stereotyped as having less innate or raw brilliance. How, precisely, is this 'stereotype' encoded into our beliefs or concept of women?

Loosely speaking, there is nothing wrong with talking about this bias as resulting from a particular stereotype of women. However, once we start seeking for deeper theoretical accounts of this bias, talk of stereotypes might lead us to seek in the wrong place.

Indeed, many discussions of socially-relevant biases—including explicit and implicit gender biases—have overlooked an important way in which biases can be encoded in conceptual structures. Most accounts of implicit bias focus on 'associations' between features and the concepts which we use to represent certain categories. Associative properties include things such as cue validity, typicality, and saliency; all properties that have been studied as part of the traditional 'prototype' theory of concepts.

However, philosophers and cognitive scientists have argued that concepts also encode dependency relations between features and dimensions. Dependency relations determine the conceptual centrality of a feature f, relative to a concept C: i.e., the degree to which other features of C depend on f. Crucially, centrality and properties like cue validity, typicality and saliency can disassociate. In this paper, we argue that dependency relations between features is one important way in which significant biases are encoded in conceptual structure. (Crucially, biases that are encoded in this way have important consequences such as being preserved in certain inferences, etc.)

We will defend this claim by presenting a series of experiments which explore the way in which (apart of; something that results in) the women have less innate brilliance gender bias could be encoded. We will argue that there is a bias against women, encoded in certain clear differences in the dependency relations between features used to represent (academic) women and men. Crucially, we will argue that this bias begins to be revealed only when we start exploring the centrality of features like SMART and HARD-WORKING, for our conceptions of certain academic class of women and men, and not when we just consider their statistical or more generally 'associative' properties within those classes.

2 Study 1

In Study 1 consists of two studies in which we directly asked people to give their estimates about the proportion of smartness and hardwork in the class of female and male professors. In both studies, each participant received only a single question, so they could not censor or anchor their responses by directly giving the same answers for the female and the male versions of the questions.

2.1 Methods of Study 1a

190 participants were assigned to one of the following questions:

(1) Consider the class of female/male professors. What percentage of all those professors do you think is very smart? Please give your best estimate.

(2) Consider the class of female/male professors. What percentage of all those professors do you think is really hardworking? Please give your best estimate.

2.2 Results

The results of Study 1a appear in Table 1. There is no significant difference in the proportion of female vs male professors who are believed to be very smart. There is a small but non-significant difference in the proportion of female vs male professors who are believed to be hardworking. Slightly more female than male professors are believed to be hardworking.

Table 1: Results Study 1a

	Male	Female
hardworking	61.3	67.8
smart	76.1	75.8

2.3 Methods of Study 1b

Study 1b is a different way of approaching the same information as that provided by Study 1a. In this case, we presented participants with an arbitrary female or male professor, and asked them to rate, on a 7-point Likert scale (1= 'not very likely' and 7 = 'quite likely'), one of the following questions:

(3) Mary/Jack is a professor. How likely do you think it is that she/he is very smart?

(4) Mary/Jack is a professor. How likely do you think it is that she/he is really hardworking?

130 participants were each assigned to one question.

2.4 Results

The results of Study 1b appear in Table 2. The numbers tell basically the same story as those obtained in Study 1a. Participants believe that a random female or male professors is quite likely to be both smart and hardworking. At the same time, there is no significant difference in those likelihood for a female compare to a male professor. Again, the small non-significant differences obtained in this case are in the same trend as those obtained in Study 1a: men get better higher numbers for smartness, and women for hardworking. Still, this differences are not significant in either of the studies.

Table 2: Results Study 1b

	Male	Female
hardworking	5.7	5.9
smart	6.0	5.9

2.5 Discussion

Studies 1a-b provide some useful background results regarding the beliefs of participants about the distribution of properties such as being smart and hardworking amongst the female and male class of professors. Although we used two different measures to get at the subjective judgments about such distributions, both measures tell basically the same story. Participants believe that both female and male professors are highly likely to be both smart and hardworking. Although in both cases there is an observable trend such that male professors are believed to be somewhat more likely to be smart compared to female professors, and female professors are believed to be somewhat more likely to be hardworking that their male counterparts, neither of these differences was significant. Studies 1a-b provide us with corroborating evidence about participant's explicit beliefs, and will therefore serve as a background to help us interpret the following Studies.

3 Study 2

In Study 2 we used a variation on a standard free production procedure to generate prototypes for both female and male professors. One aim of Study 2 is to see how the prototype freely generated by subjects compares to the direct estimates observed in Study 1a-b. The resulting differences/ similarities might begin to inform us about the way in which biases could be encoded in conceptual structure.

3.1 Methods

99 participants were recruited through Amazon Mechanical Turk and reimbursed for their participation. Participants were randomly assigned to either the condition featuring the female professor Mary (N=43) or the male professor Jack (N=46). Four participants were excluded because they were not native English speakers. The stimuli read as follows:

(5) Imagine that Mary/Jack is a professor at a university. Please list five features that you think are typical of Mary/Jack.

The participants were then provided with five blank rows in which they could list the five features.

3.2 Results

Table 3 shows the percentage of participants who listed, in each of the female and male condition, *hardworking* or its synonyms (e.g., *dedicated*) among the five features. Figure 1 also shows the percentage of participants who listed, in each of the female and male condition, *smart* or its synonyms (e.g., *brilliant*, *intelligent*). 'Smartness' was the feature most frequently produced in both the female and male condition. Furthermore, 'hardworking' was the second feature most frequently produced in both the female and male condition. Now, the production frequency of 'smartness' features was slightly (but not significantly higher) in female (76.6%) than in the male condition (72.4%). However, the production frequency of 'hardworking' features was almost twice as high in the female (39.5%) than in the male (21.3%) condition. This difference is highly significant.

Table 3: Results Study 2 (in terms of percentage of participants)

	Male	Female
hardworking	21%	40%
smart	72 %	77%

3.3 Discussion

Study 2 has both remarkable similarities and differences when compared with Studies 1a-b. Again, we observe that both female and male professors are believed to be characteristically smart. In this task, participants were free to produce whatever features came to mind, and because it was a between subjects design, no participants could anchor their responses to a previous response involving an exemplar of the opposite sex. Still, the story with respect to smartness and the class of female and male professors is consistent with our previous result, suggestion, again, that participants do not believe that there is any gender difference, within this class. However, we now observe a new results. Namely, the prototype freely generated for female professor has a much higher weight (as a proportion of production frequency) in the feature 'hardwork' compared to that generated for a male professor.

How can we make sense of the difference between Studies 1a-b and Study 2, with respect to the role of hardwork amongst the class of female and male professors? Note that Studies 1a-b asked participants about statistical properties, whereas Study 2 asked them to freely produce typical or characteristic properties. Experimenters know that frequency in the class is often an important determinant of typicality, in other words, is often a reason why a feature is considered part of a prototype. However, this is not generally the case. For example, if we ask participants what proportion of lions they think have stomachs, they will probably answer that all or most of them. At the same time, 'having stomachs' is rarely produced as typical or characteristic feature of lions. This is because it has a very low diagnostic value.

Since we already have (based on independent Studies, which used different measures) a good idea about what the frequency of 'hardwork' is, according to participants, for both female and male professors (namely, about the same), we can reasonably conclude that we need a different explanation for why 'hardwork' is considered a more typical or characteristic feature of women compared to men. Now, features might obtain high typicality as a function of various properties—diagnosticity, frequency, saliency etc.—so we need a different experimental paradigm.

4 Study 3

Study 2 suggests that one important difference between the concept of, or the way in which people encode information about, female and male professors is that the feature of being hardworking is significantly more important, characteristic, or typical for female than for male professors. Now, because prototypes can encode features with various properties—including cue validity, saliency etc.—Study 2 does not, by itself, tell us why that feature might be differentially encoded. It follows that Study 2 does not yet tell us how that clear gender difference could be related to the sort of less brilliant bias which Leslie et al. (2015), and indeed many other people assume exists against women.

At this point, the notion of centrality can guide our empirical investigation. In particular, we argued at the outset that important biases could be encoded, not in (the subjection judgments of) the statistical (or perceptual, such as saliency) properties of features used to represent classes, but rather in the network of dependencies that connect these features. For example, the concept of a 'breakfast chair' and an 'office chair' might both have the feature HAS A BACK. However, the centrality of that feature might differ because believe that office chairs really do need a back (since you seat on them for many hours), whereas formal dinning chairs could do their job quite well even if they do not have a back (since you just use them to eat breakfast, which is not that long). This belief can be austerely encoded by simply having more features of office chair depend on having a back, than do features of breakfast chair.

One way in which Frank Keil has suggested that we can get at the dependency structure of concepts is by asking people to generate features, as in the free production task, but in response to simple reasoning schemes. We adapted with basic idea, and presented subjected with a causally guided reasoning scheme. Given our aim, we wanted to see what feature of arbitrary female and male professors participants intuitively think 'explains' their having become a professor.

How does this help us understand the result that, according to the free production task in Study 2, hardwork is more typical/characteristic/important for the representation of female compared to male professors? If the production frequency remains as before, this would strongly suggest that the reason why women are thought to be more hardworking is because they need to be so more than men in order to become professors. If the difference is not preserved, this suggests that the difference in the weight of hardwork for the concept of a female vs male professor is independent of its relative centrality in each concept.

4.1 Methods

87 participants were randomly assigned to either one of the condition featuring the female professor (N=36) or the male professor Jack (N=51). The respective reasoning schemes had the following form:

> Becoming a professor is hard (difficult?).
> Mary/Jack has recently become a professor.
> Therefore, Mary/Jack must be ____________________

Participants were then asked to enter the feature that they think would best fit the reasoning scheme.

4.2 Results

In both the female and the male condition, terms related to commitment and intelligence made up for over 50% of all responses. As in the case of the free production task, no difference was found, between those conditions, in the production frequency of 'smartness' and its synonyms: 27.8% of participants produced that feature in the female condition, and 27.5% produced in the male condition. However, crucially, the production frequency of 'hardwork' and its synonyms was significantly different in the female vs the male condition: 38.9% of participants cited it in the female condition and 27.5% cited it in the male condition.

Table 4: Results Study 3 (in terms of percentage of participants)

	Male	Female
hardworking	27.5%	40%
smart	27.8 %	27.5%

4.3 Discussion

The aim of Study 3 was to channel participants to freely produce features, in response to a very simple reasoning scheme. This allows us to get some insights into the properties of these features, and why they might be encoded in the representations for, or concepts of, female and male professors. Due to the kind of reasoning scheme we used, we only asked subjects for a single feature that makes most sense of the scheme. Therefore, we cannot draw any when comparing the absolute numbers of this study and Study 2, in which subjects could cite up to five features. However, we can draw some insights by comparing the similarities and differences between the overall structural pattern of the productions frequencies for hardwork and smart in the female and male condition.

In the case of male professors, hardwork and intelligence are about equally important/frequent character traits that explain why an individual could become a professor. In the case of female professors, hardwork is a significantly more important/frequent character trait than intelligence to explain why an individual could become a professor. Now, because the stimuli did assert that becoming a professor is difficult, this could have increased the absolute numbers for hardwork and its synonyms (we control for this in a future experiment) via some form of priming. However, this effect should apply both to the female and the male condition, hence it cannot explain the huge difference in the results.

Most importantly, this strongly suggests that the reason why 'hardwork' has a higher weight/importance for female compared to male professors is because participants believe (implicitly or explicitly) that women have to be more hardworking than men to become professors. Is this because they face more obstacles? Or is it because they have less raw talent, and therefore need to work harder?

References

Leslie, S.-J., A. Cimpian, M. Meyer, and E. Freeland (2015). Expectations of brilliance underlie gender distribution across hte academic disciplines. *Science 347*(6219), 262–265.