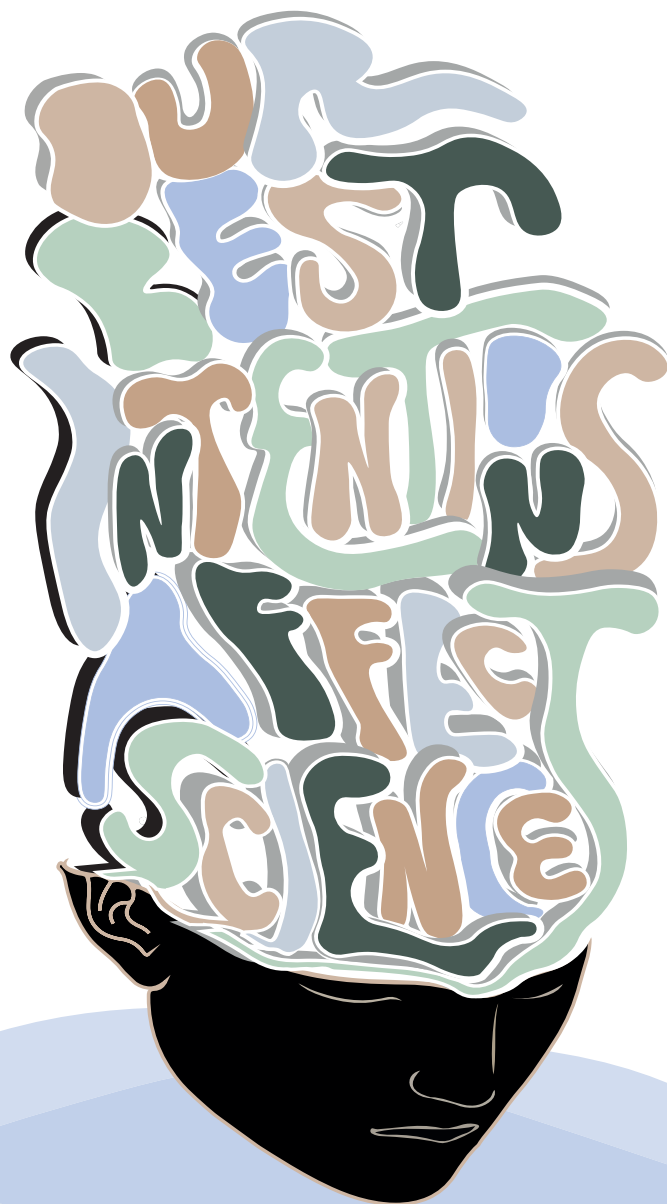


# JOURNAL OF NEUROSCIENCE AND COGNITION



Internal report of the Master Neuroscience and Cognition

Introducing  
Bias

Pinpointing  
Bias

Shifting  
Focus

# Table of Contents

## SECTION 1

### INTRODUCING BIAS

Wired to be biased	6
Definitions of the human being	8
Cyclops: A myth based on truth	12

## SECTION 2

### PINPOINTING BIAS

Bias in Social Neuroscience	16
Bias in GWAS	20
The Little Things Matter	25
Language & Schizophrenia	29
Disorders in Context	36
Brain. The Final Frontier?	39
South Africa and Hormone Research	40
Introducing Elisabeth Bik	42
PhD Perspective: Lab Culture	48
PhD Perspective: USA & the Netherlands	49

Dear reader,

Before you lies the first issue of the 16th volume of the Journal of Neuroscience and Cognition. The Board has succeeded in putting together a very interesting and relevant issue of the Journal of Neuroscience and Cognition. The theme of this journal is "BIAS: our Best Intentions Affect Science". For your convenience the journal has been divided into subsections, starting off with a section in which the concept bias is introduced, moving on to a section in which the concept bias is explored in more detail, and ending with a section on future challenges with respect to the concept bias.

The theme bias is relevant to science in the broadest sense. Biases in methodology and analyses techniques may impact the way in which we interpret our research data. You can read more about this in for example a student piece by Sanne Rodenburg on the amygdalocentric focus in fear conditioning research, or an interview with David Terburg on biases in social neuroscience. Biases can also affect clinical practice, for example the way in which we label illnesses may impact the way in which we think about and discuss those illnesses. This is further illustrated with respect to psychiatry in a piece that focuses on the way in which our language shapes our understanding of psychiatric illnesses. Furthermore, our society in general, and science specifically, needs to deal with biases in how, for example, students and employees are selected for certain positions. Anyone who has ever set foot in the "Senaatszaal" in the Academy Building of Utrecht University will understand exactly what I mean. To conclude, we should not only focus on minimising bias in the present, we should also make sure to avoid bias in the future. You can read more on this challenge within a piece by Julia Berezutskaya on bias in AI and machine learning.

The Editorial Board has done a fantastic job in selecting interesting and inspiring content. I also would like to thank all the contributors for freeing up time in their schedule to write a piece for the journal or be interviewed.

Anouk Keizer  
Senior supervisor



You can find all of the **references** on our website!

## SECTION 3

### SHIFTING FOCUS

Patient Integration into Research	52
Let Them Hear Your Voice!	55
Hidden Biases in Techniques	56
Bias in Neuroscience	59
Bias in Artificial Intelligence	61
Amygdala and Fear Research	63
Classification of Psychiatric Disorders	66
Scientific Integrity Game	70
Neuroscience in the Media	74
The Binary Brain: Myth or Bust	76
The Fundamentals of Neuroscience	78
Diversity Poll	80
Our Board	80
Our Contributors	81

Dear reader,

We, the board of the Journal Neuroscience and Cognition, are happy to present you the first issue of 2022. For this issue, we have chosen the topic of bias. While brainstorming for a theme, we kept coming back to the concept of bias and how it, unfortunately, still plays a large role within research. Furthermore, we noticed how important this topic is and enjoyed being able to learn more about it from various points of views.

We wanted to make this topic visible from various perspectives, by both pinpointing biases that can occur within the whole scientific community, such as bias in funding, as well as biases that occur in specific fields. By doing so, we hope to shed light on these topics and start a discussion, as we believe that we can work to counteract them if we are aware of our own biases and those occurring around us.

Of course, people became aware of how bias can influence science long before us. There has been incredible work done to reduce biases, such as the implementation of new techniques, or giving patients a voice to help in research. However, even though many are working hard in shifting the focus, it is often a slow process, as Floortje Scheepers and Fleur Velders mention in their piece about the dismantlement of the current method of categorizing mental disorders. By presenting you with this variety of options and ideas of how to counteract our biases to improve research, we hope to make you aware of the biases present in your field and spark your interest to look into ways how to reduce them.

In addition to all the wonderful articles, we also want to highlight the artwork we received from students who were inspired by our theme.

We hope you will enjoy reading this issue of the journal!

Yours sincerely,

Chiara Galizia

Editor in Chief

*\* There is no conflict of interest to disclose.*

# BIAS IN ARTIFICIAL INTELLIGENCE

DR. JULIA BEREZUTSKAYA<sup>1</sup>

<sup>1</sup> *DONDERS CENTER FOR BRAIN, COGNITION & BEHAVIOR, ARTIFICIAL INTELLIGENCE DEPARTMENT, RADBOUD UNIVERSITY, NIJMEGEN, THE NETHERLANDS*

Machine learning and artificial intelligence (AI) methods have come to dominate many fields, including computer vision, natural language processing, robotics, healthcare and finances. They are being used to build highly accurate solutions for real-world problems. AI models benefit from large datasets, which can make them more robust against noisy and missing data compared to conventional statistical models. Among some examples of applications are medical image analysis (clinical diagnosis based on medical imaging), machine translation between languages (Google Translate), self-driving cars, research on brain-computer interfaces, and many others.

Despite multiple advantages that AI models provide, there are also issues that raise concern. One of such concerns is an issue of bias in AI. Biases in an AI model represent trends of the system: something that the model has learned and leans towards based on the data it has seen or the algorithm it has been trained with. A famous example includes gender bias in AI language models (Floridi & Chiriatti, 2020), that are trained on millions of language texts. Specifically, when a trained AI language model was asked to complete a sentence “She was a competent ...” it would propose words such as “waitress”, “nurse”, “receptionist” (female leaning occupations), whereas the sentence “He was a competent ...” would be completed with “boss”, “professor”, “banker” (male leaning occupations). Such language models have been shown to reflect stereotypes and biases, based on gender, race, religion, age and other characteristics (Brown et al., 2020). Another example is a scrapped Amazon e-recruitment AI system based on automatic processing of resumes, that ended up hiring a disproportionate number of male candidates (Hunkenschroer & Luetge, 2022). Moreover, the very fact that most voice assistants have female voices (Siri, Cortana, Alexa, Bixby) has been seen as further reinforcement of gender bias in AI, as they encourage stereotypes of women as submissive and compliant (Chin & Robison, 2020).

Such bias is not specific to text-based AI systems or social discrimination for that matter. In healthcare, suboptimal handling of input medical images or model training mechanisms can lead to biased predictions in disease diagnosis. Specifically, it has been suggested that AI models tend to underestimate the risks of false negatives (i.e. missing a diagnosis) (Megler & Gregoire, 2019) as analysing medical data is merely an image classification task for them. Moreover, these systems can lack a wider context of the real-world problem that includes external factors and special cases, and can therefore misrepresent individual patient characteristics. Some examples of this include incorrect cardiovascular disease assessment in black individuals due to overrepresentation of white non-Hispanic patients in the training data (Tat et al., 2020), and poor performance of cancer prediction systems trained on small datasets that misrepresent complex variation across patients that naturally occur in real data (Kleppe et al., 2021). Such statistical bias in AI-based decision making can clearly result in wrong outcomes, and increased risks for individual patients.

Researchers in machine learning and AI spend a lot of effort trying to understand the sources of bias, and correct them to create better models. Often the bias can be mitigated by collecting more data of the underrepresented type and thereby making the dataset more balanced, using better training schemes, such as adding a sensitivity component that guides model learning (Serna et al., 2022), or more advanced algorithms (Pena et al., 2020) that explicitly try to anticipate and correct known issues. Importantly, we need to continue creating more bias awareness by openly discussing it within our society, reporting cases of AI-based discrimination and inequality, and promoting diversity and equality in our society overall. These efforts can encourage extensive checks of AI systems prior to their practical use, and stimulate further research on bias analysis and correction.

Bias in and of itself does not have to be a negative phenomenon. In fact, the goal of an AI system is to establish complex relationships between inputs (e.g. medical images) and outputs (diagnosis) by learning some hidden associations and useful trends in the data that help make correct predictions. And even though many AI systems continue to remain “black-boxes” (which means we do not really know how these models make predictions), there has been an effort to develop more interpretable, transparent systems that could be constrained (i.e. biased) by useful assumptions through learning. For example, latest AI models can develop an in-

ductive bias for relational and hierarchical structures in the data (Hamrick et al., 2018), can incorporate expert knowledge (Omlin & Snyders, 2003) or learn to mimic aspects of human reasoning, such as analogy thinking and use of heuristics (Goyal & Bengio, 2020), which can all be helpful in solving a complex real-world problem. Our hope is that the AI field will learn to deal with harmful biases more and more efficiently, and capitalise on good, useful trends for solving multiple tasks, thereby establishing a new standard for responsible AI.

***“Bias in and of itself does not have to be a negative phenomenon.”***





The content of this journal must not be reproduced in any form and to any extent without written permission of the Editor in Chief and Senior Advisor. This journal is meant only and specifically for the students of the Master 'Neuroscience and Cognition' of Utrecht University.