

Administrative	
<b>Project Name</b>	Open Science WS2022: Reproduction of a Study
<b>Project Description</b>	The purpose of this work is to attempt to reproduce a study by re-using a published dataset, for the final project of the TH Köln ( <a href="https://www.th-koeln.de/">https://www.th-koeln.de/</a> ) module "Open Science" as part of the Digital Sciences Master's Degree. This module is taught by Prof. Dr. Mirjam Blümm ( <a href="mailto:mirjam.bluemm@th-koeln.de">mirjam.bluemm@th-koeln.de</a> ) and Prof. Dr. Claudia Frick ( <a href="mailto:claudia.frick@th-koeln.de">claudia.frick@th-koeln.de</a> ).
<b>Researcher Names</b>	Natasha Randall and Berrak Küçük.
<b>Researcher Contacts</b>	<a href="mailto:natasha_jacqueline.randall@smail.th-koeln.de">natasha_jacqueline.randall@smail.th-koeln.de</a> <a href="mailto:berrak.kuecuek@smail.th-koeln.de">berrak.kuecuek@smail.th-koeln.de</a>
<b>Responsibilities</b>	Both of the researchers are responsible for implementing the DMP, maintaining the datasets, and all stages in the research lifecycle.
<b>Funding</b>	No funding or competing interests.
<b>Date of Last Update</b>	02/01/2023.
<b>Related policies</b>	<ul style="list-style-type: none"> <li>The licence of the original dataset is: "CC0 1.0 Universal (CC0 1.0) Public Domain Dedication". Therefore we may modify and redistribute the data freely. (<a href="https://creativecommons.org/publicdomain/zero/1.0/">https://creativecommons.org/publicdomain/zero/1.0/</a>)</li> <li>TH Köln "supports the Open Access initiative" and recommends researchers "take principles of Open Science into account", however provides no hard requirements. (<a href="https://www.th-koeln.de/mam/downloads/deutsch/hochschule/open_science_policy.pdf">https://www.th-koeln.de/mam/downloads/deutsch/hochschule/open_science_policy.pdf</a>)</li> <li>We aim to follow the Open Research Europe Data Guidelines (<a href="https://open-research-europe.ec.europa.eu/for-authors/data-guidelines">https://open-research-europe.ec.europa.eu/for-authors/data-guidelines</a>) including making data FAIR, and following the standards for spreadsheet data and software distribution.</li> </ul>
Data	
<b>Provenance</b>	The dataset was collected for the study "Women's preference for masculine traits is disrupted by images of male-on-female aggression" ( <a href="https://doi.org/10.1371/journal.pone.0110497">https://doi.org/10.1371/journal.pone.0110497</a> ). It is available on DRYAD ( <a href="https://datadryad.org/stash/dataset/doi:10.5061/dryad.9bg43">https://datadryad.org/stash/dataset/doi:10.5061/dryad.9bg43</a> ).
<b>Legality</b>	<ul style="list-style-type: none"> <li>The provided dataset is originally licenced under CC0 1.0 Universal (CC0 1.0) Public Domain Dedication.</li> <li>There is no licence given for the provided images, therefore they will not be re-released.</li> <li>Our modified dataset is released under Creative Commons Attribution 4.0 International (CC BY 4.0).</li> </ul> <p>Our licence choices are made in order to promote open practices and encourage re-use of the data.</p>
<b>Format</b>	The dataset is in the form of a tabular spreadsheet. The original data is stored as an .xlsx file. Our modified datasets are stored as .csv files, due to ease of interoperability and its standard as a format. The size of the data is very small: the entire project is less than 50MB. The supplied images are of format .BMP.
<b>Ethics</b>	As there is no available licence or provided guidelines for the sharing of the participants' photos, we choose to view these data as sensitive, personal data. Following the standard "as open as possible, as closed as necessary", the photos will therefore not be released.

Storage	<ul style="list-style-type: none"><li>• The data is stored both locally on the researchers' personal machines, and available online on GitHub.</li><li>• Although GitHub has some problems, it was chosen due to its accessibility: the code in the Jupyter notebooks can be explored online without the reader needing to install any software, and the associated Readme file and documentation aids in understandability. GitHub also provides automatic version control.</li><li>• The personal data (photos of participants) are secured in an encrypted folder on the researchers' personal machines, and not made available on Github.</li></ul>	
Preservation	We are both responsible for preserving the data. In order to ensure long term preservation, the data is stored in multiple locations, both locally and in the cloud. Regular backups are also made of the data.	
FAIR	<p><b>Findable:</b> The data is available and findable on GitHub, and linked in the Readme file to the DOI identifiers of both the original dataset and study. The GitHub repository is tagged with relevant labels to encourage findability.</p> <p><b>Accessible:</b> The Jupyter notebooks and datasets can be accessed and read directly on GitHub, improving accessibility. For local use, the datasets and code files can be downloaded, and do not require any proprietary software to access.</p> <p><b>Interoperable:</b> The format of the files use common and open standards, e.g. .csv. The language of the software is Python and R, both commonly used languages for scientific computing.</p> <p><b>Reusable:</b> A detailed data documentation and metadata description are provided, describing both the directory structure and contents of the dataset, to aid in reusability. Data usage licences are provided.</p>	
Metadata		
Project Directory Structure	The project directory contains two main folders:	
	“Original Files”	The original study files as provided by the original researchers.
	“Reproduction Project”	Our re-structuring and modifying of the original files, as well as additional datasets and software (code).
	The “Reproduction Project” folder contains:	
	4 folders corresponding to the 2 “main studies” and 2 “supplementary studies” described in the paper	Each folder contains its respective data, e.g. “main_study_face_data.csv” and any re-used and modified datasets, e.g. “main_study_face_data_modified.csv”.
	The folder “Reproduction Project Code”	Containing the code and Jupyter notebooks used to analyse the data, numbered in order. The notebooks containing an incorrect analysis (labelled “WRONG DATA”) are included for transparency.
	The xlsx file: “Downloaded Data (All Studies)”.	The form of the original dataset, included for reference.
	The pdf file “Published Article”.	The original journal article associated with the dataset, included for reference.
	The docx file “Description of Supplementary Studies”.	An explanation of the supplementary studies as provided by the original researchers, included for reference.

<b>Data Documentation</b>	<p>The purpose of the original data collection, structure of the associated study, and meaning of the attributes in the dataset is described in the journal article: (<a href="https://doi.org/10.1371/journal.pone.0110497">https://doi.org/10.1371/journal.pone.0110497</a>).</p> <p>In our modified dataset, ("main_study_face_data_cleaned.csv") the contents of each of the attributes are:</p> <table><tr><th>Attribute Name</th><th>Attribute Meaning</th><th>Attribute Contents</th></tr><tr><td>participant_id</td><td>The unique ID identifier for each of the participants in the study.</td><td>String of 331 possible values, e.g. "13" or "211b".</td></tr><tr><td>prime_condition</td><td>A number representing the priming condition group each participant was assigned to.</td><td>Integer in range: 1 to 5</td></tr><tr><td>prime_condition_names</td><td>The name (mapped from the prime condition number) of the relevant priming condition group.</td><td>String of possible values: "neutral", "male/male", "male/group", "male/female", "pathogen"</td></tr><tr><td>trial_number</td><td>Each row in the dataset corresponds to each trial for each particular participant, where they choose whether they prefer an image of a masculinised or feminised face.</td><td>Integer in range: 1 to 40</td></tr><tr><td>image</td><td>The image shown to the participant, containing a masculinised and feminised face.</td><td>String of form Slide{}.bmp, where {} is the number of the slide, e.g. Slide13.bmp</td></tr><tr><td>pre_post_prime</td><td>An indicator of whether this particular trial occurs before (0) or after (1) the participant has been shown their respective priming images. All trials <math>\geq 20</math> are post prime (1), the remainder are pre-prime (0).</td><td>Integers: 0 or 1</td></tr><tr><td>chose_masc</td><td>An indicator of whether on this particular trial the participant chose the masculinised face (1) or feminised face (0).</td><td>Integers: 0 or 1</td></tr></table>	Attribute Name	Attribute Meaning	Attribute Contents	participant_id	The unique ID identifier for each of the participants in the study.	String of 331 possible values, e.g. "13" or "211b".	prime_condition	A number representing the priming condition group each participant was assigned to.	Integer in range: 1 to 5	prime_condition_names	The name (mapped from the prime condition number) of the relevant priming condition group.	String of possible values: "neutral", "male/male", "male/group", "male/female", "pathogen"	trial_number	Each row in the dataset corresponds to each trial for each particular participant, where they choose whether they prefer an image of a masculinised or feminised face.	Integer in range: 1 to 40	image	The image shown to the participant, containing a masculinised and feminised face.	String of form Slide{}.bmp, where {} is the number of the slide, e.g. Slide13.bmp	pre_post_prime	An indicator of whether this particular trial occurs before (0) or after (1) the participant has been shown their respective priming images. All trials $\geq 20$ are post prime (1), the remainder are pre-prime (0).	Integers: 0 or 1	chose_masc	An indicator of whether on this particular trial the participant chose the masculinised face (1) or feminised face (0).	Integers: 0 or 1
Attribute Name	Attribute Meaning	Attribute Contents																							
participant_id	The unique ID identifier for each of the participants in the study.	String of 331 possible values, e.g. "13" or "211b".																							
prime_condition	A number representing the priming condition group each participant was assigned to.	Integer in range: 1 to 5																							
prime_condition_names	The name (mapped from the prime condition number) of the relevant priming condition group.	String of possible values: "neutral", "male/male", "male/group", "male/female", "pathogen"																							
trial_number	Each row in the dataset corresponds to each trial for each particular participant, where they choose whether they prefer an image of a masculinised or feminised face.	Integer in range: 1 to 40																							
image	The image shown to the participant, containing a masculinised and feminised face.	String of form Slide{}.bmp, where {} is the number of the slide, e.g. Slide13.bmp																							
pre_post_prime	An indicator of whether this particular trial occurs before (0) or after (1) the participant has been shown their respective priming images. All trials $\geq 20$ are post prime (1), the remainder are pre-prime (0).	Integers: 0 or 1																							
chose_masc	An indicator of whether on this particular trial the participant chose the masculinised face (1) or feminised face (0).	Integers: 0 or 1																							

Related Software	
<b>Research Software</b>	In order to encourage re-use and transparency, all of the code relating to the modification and analysis of the dataset is also available. The code is stored in Jupyter notebooks alongside full documentation and comments. The results can therefore be viewed and accessed even without needing to run the code.
<b>Software Requirements</b>	In order to run the code, a Python and R distribution are required. The Jupyter Notebooks can be viewed on Github with no installation necessary.
<b>Software Licence</b>	Our research software is released under the MIT licence. Our licence choices are made in order to promote open practices and encourage re-use of the code.