

Article

Not peer-reviewed version

From “Bob/Kirk” to “Błażej/Piotrek” Effect – Replication of Social Bias Effect Towards People Whose Names Match Their Faces

[Kamil Stawiany](#)^{*}, Krzysztof Kozak, Zuzanna Sokołowska

Posted Date: 2 January 2025

doi: 10.20944/preprints202501.0057.v1

Keywords: Sympathy; name-face pairing; coherence; cross-modal perception; Barton and Halberstadt effect; replication; experiment; SONA platform



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Article

From “Bob/Kirk” to “Błażej/Piotrek” Effect— Replication of Social Bias Effect Towards People Whose Names Match Their Faces

Kamil Stawiany *, Krzysztof Kozak and Zuzanna Sokołowska

Poland; kstawiany@st.swps.edu.pl

Abstract: The study explores the hypothesis that individuals show greater sympathy towards coherent name-face pairs compared to incoherent pairs, replicating the effect observed by Barton and Halberstadt. A structured questionnaire was designed where participants rated their sympathy for each name-face pair on a scale from 1 to 9, with 1 signifying “Would not like at all” and 9 denoting “Would like very much.” Recruitment was conducted through the university’s SONA platform, ensuring voluntary participation, anonymity, and confidentiality. Names and faces were paired based on responses from a pilot study, creating highly coherent or highly incoherent pairs. The first study was conducted with psychology students but failed to replicate the previous effect. As a result, a second experiment was carried out using the same survey with a general population, excluding psychology students. However, the effect still could not be replicated, raising questions about the extent to which cross-modal perception is as prevalent as previously thought.

Keywords: sympathy; name-face pairing; coherence; cross-modal perception; barton and halberstadt effect; replication; experiment; SONA platform

Introduction

Everything began in 1924 when Usnadze (1924) asked people to assign abstract names to abstract shapes. Five years later, Köhler (1929) designed the now well-known shapes and named them ‘Takete’ and ‘Maluma.’ He asked people to assign these names to two shapes he created. One rounded and the other angular. It turned out that people more often associated the word ‘Takete’ with the sharp shape and ‘Maluma’ with the rounded one. When the words ‘Kiki’ and ‘Bouba’ appeared, a ‘modern’ approach to studying intermodal perception emerged. Research has shown that 71.1% of people associate the sharp shape with the word ‘Kiki’ and the rounded shape with the word ‘Bouba,’ even though the study sample spoke 25 different languages (Ćwiek et al., 2021). When we talk about names, Barton & Halberstadt (2022) say that it is possible that associations between abstract concepts and their names may be related to the shape of the mouth as it forms the sounds and the sound waves produced. We tried to replicate this study in our article as well. Also we would like to mention that currently science faces a replicability problem and only articles that showed an effect will be published making the false positive (Schmidt & Oh, 2016). Also to make replications the best as they can be they must be performed all around the world across a variety of cultures (Milfont & Klein, 2018).

Pilot Study

This pilot study investigates the relationship between the perception of angularity and roundness in names and corresponding facial images among 30 participants (19 females, 11 males). We establish a database of 40 names categorized as ‘angular’ or ‘round’ from the 100 most popular Polish names (Ministerstwo Cyfryzacji, 2020), and generate 40 facial images using an AI algorithm, categorized into 20 angular and 20 round faces using biometric tools (Farkas, 1994). Participants were

asked to rate each name and face on a scale of 1 to 9, where 1 corresponds to a 'BOUBA' shape (round) and 9 to a 'KIKI' shape (angular). The images of shapes are present on the left and right of the scale to help participants with associating faces and names to shapes. The results help identify coherent and non-coherent name-face correlations, used in the first and second study to compare these groups and test the hypothesis. It is worth mentioning that we used only the faces of white men to minimize gender bias (Barton & Halberstadt, 2018). The angularity-roundness will be determined by method provided by Farkas (2022) and for the names we use a similar method provided by Sidhu, Pexman & Saint-Aubin (2016) where "I", "E", "P", "T", "K" will be labeled as 3 points, "A" 2 points and "L", "M", "N", "B", "D", "U", "O" will be considered as 1 point. Then we will sum all the corresponding points and divide by the number of counted letters. If a letter does not have any point assigned (like the letter "R"), we will skip it.

Method

In the initial phase of the analysis, descriptive statistics were computed to summarize key factors, including the mean, median, standard deviation, skewness, kurtosis, and range (minimum and maximum values) for each variable. The normality of the data distribution was assessed using the Shapiro-Wilk test, with results indicating non-normal distribution across several variables. Consequently, Spearman's Rho correlations were used for further analysis.

Cronbach's α was calculated to evaluate the reliability of each factor, including angular and round names as well as angular and round faces, with results compiled for each factor. Higher reliability for rounded faces was observed, leading to insights regarding the measurement tool's adequacy for identifying rounded versus angular faces. Reliability for names was also assessed, where specific phoneme exclusions were noted for their potential impact on ratings.

For participant selection, a target sample size of 272 was initially set, and actual participation reached 420 respondents in Study 1, which included psychology students. Descriptive statistics and statistical tests were computed for consistent and inconsistent face-name pairings. The Wilcoxon rank-sum test was applied to compare these groups due to non-normal data distribution.

In Study 2, data collection was extended to a general population sample of 323 respondents with exclusions to avoid psychology student bias. The same statistical analyses were repeated, and descriptive statistics for each group were calculated, with the Wilcoxon rank-sum test again used for group comparison.

Results

In the first step of the analysis, basic descriptive statistics were calculated along with an analysis of the normal distribution. Detailed analysis results are presented in Table 1.

Table 1. Factor analysis.

Factor	M	Mdn	SD	Sk.	Kurt.	Min.	Maks.	S-W	p
Angular names	5,59	5,82	0,96	-1,26	2,88	2,40	7,10	0,92	0,022
Round names	4,44	4,65	0,84	-1,03	0,38	2,40	5,55	0,91	0,011
Angular faces	5,34	5,21	0,76	0,21	-0,46	3,79	6,74	0,97	0,506
Round faces	3,72	3,76	1,25	1,09	2,56	1,38	7,76	0,93	0,043

M – mean, Mdn – median, SD – standard deviation, Sk. – skewness, Kurt. – kurtosis, Min. – minimum value, Max. – maximum value, S-W – Shapiro-Wilk test, p – statistical significance of the S-W test.

The Shapiro-Wilk test showed that the data do not follow a normal distribution. Therefore, Spearman’s Rho correlations were calculated, and five pairs of name-face combinations were selected that correlated the most strongly, both positively (consistent) and negatively (inconsistent). Rounded faces demonstrated higher reliability than sharp faces, indicating that this tool allows us to predict which rounded faces will appear rounded to the participants, but we cannot say the same for sharp faces with the same confidence. Regarding names, sharp names demonstrated higher reliability, likely due to the fact that we decided to omit the phoneme ‘R,’ which often appears as sharp in Polish but rounded in English (Winter et al., 2022), thus disrupting the effectiveness of our tool. The average rating for sharp names was 5.59, with a median of 5.82 and a standard deviation of 0.96. Ratings for rounded names had an average of 4.44, a median of 4.65, and a standard deviation of 0.84. To ensure that the statistical inference was based on reliable tools, the reliability of the variables was checked using Cronbach’s α test. The reliability results for each factor are presented in Table 2.

Table 2. Reliability analysis results for each factor.

Factor	Number of items	Cronbach’s α
Angular names	20	0.80
Round names	20	0,68
Angular faces	20	0,67
Round faces	20	0,92

In the results concerning face sets, the rounded faces factor shows higher reliability than the sharp faces factor. This may indicate a flaw in the tool used to select the test items. Objective criteria, such as the angle between the ears and the center of the jaw, proved sufficiently reliable in identifying rounded faces. However, the reliability analysis results suggest that the same measurement tool is not sufficiently reliable for identifying sharp faces. These findings suggest the need to supplement the tool with additional anthropometric criteria for face assessment, which would consider more aspects of facial structure.

In the case of names, the sharp names factor showed higher reliability than the rounded names factor. During the analysis, it was noted that the names Roman, Gracjan, Dorian, Kondrat, Adam, Radosław, and Brajan negatively correlated with the overall result. A common characteristic of most of these names is the presence of the phoneme ‘R.’ The reason for these results may be the omission of this frequently occurring phoneme in the Polish language when calculating the objective sharpness indicators for names. Some studies (Winter et al., 2022) suggest that this sound is perceived as sharp (rough), so it is possible that the presence of this phoneme in names might lead participants to more frequently associate the name with the shape ‘kiki,’ thereby inflating the sharpness scores of these names. Despite all of that we state that those results confirm hypotheses that objectively angular Polish names will be perceived as angular and objectively round as round.

Study 1 (Psychology Students)

The target sample size for the study is 272 participants. Age of participants was from 18 to 80 years ($M = 25.9$, $SD = 8,47$). According to Barton & Halberstadt (2018) the effect was strong (partial $\eta^2 = .40$). G-Power analysis of T-test for two independent groups ($d=0.4$; $\alpha=0.05$; $power=0.95$) suggested a sample size of 272 participants (Faul, Erdfelder, et al., 2007). But the plans differed from reality a bit. We haven’t seen so many responses so fast ever in our life. It was our first study made with the SONA system. Overwhelmed by the quantity of responses we let them flow, because we wanted to find smaller effects. We were not expecting that we had to exclude more than 100 responses so we

only considered 420. Aside from our main topic, we learned that obligatory systems that force students to generate responses can do more harm than good to the quality of data. Table 3 shows how our participants ranked each pair of name and face divided by gender. The study was preregistered on Open Science Network and all materials and data we collected are available here (Stawiany et al., 2024).

Table 3. Descriptive statistics for psychology students.

Statistics	Kind	Value
Mean	Consistent	4.91
	Inconsistent	4.86
Median	Consistent	5.00
	Inconsistent	5.00
Standard deviation	Consistent	1.89
	Inconsistent	1.97
Kurtosis	Consistent	2.42
	Inconsistent	2.38
Skewness	Consistent	-0.09
	Inconsistent	-0.05

The consistent group had a mean score of 4.91, with a median of 5.0 and a standard deviation of 1.89. The inconsistent group had a mean score of 4.86, with a median of 5.0 and a standard deviation of 1.97. The analysis results indicate that both groups have similar descriptive statistics. The mean score for the consistent group was 4.91, with a median of 5.0 and a standard deviation of 1.89, while the inconsistent group had a mean score of 4.86, a median of 5.0, and a standard deviation of 1.97. The Wilcoxon rank-sum test (conducted due to lack of normal distribution of data) for independent groups showed that the difference between the mean scores of the consistent and inconsistent groups is not statistically significant ($W = 2167294$), ($p = 0.33$), which suggests no significant difference between the groups and, consequently, no confirmation of the hypothesis which stated that consistent face-name pairs will be viewed more favorably than inconsistent pairs. SONA system turned out to be problematic. Due to the obligation of completing surveys by students to pass their bachelors program, the quality of data was questionable. Some students said they have only primary education or that they have 100 years. Some of them didn't agree to participate in the survey only to end it quickly and get points. Of course those answers were discarded and not included in our analysis.

Study 2 (General Population)

We will admit that we didn't expect the effect to not replicate at all. So we added a preregistration note that states we will run another study with the same exclusion criteria as the previous one but with the addition of education level and size of the city that participants live in. Also, we excluded people that are psychology students, because we speculate that being a psychology student can lead to bias while doing our survey (Stawiany et al., 2024).

This time, our considered sample size was 323 respondents. Their age ranged from 18 to 72 years ($M = 27.3$, $SD = 8.42$). This time we didn't have as many problems with attention checks as with

incomplete surveys. This time we removed more than 150 survey responses. Table 4 shows descriptive statistics for considered sample size.

Table 4. Descriptive statistics for general population.

Statistics	Kind	Value
Mean	Consistent	5.30
	Inconsistent	5.23
Median	Consistent	5.00
	Inconsistent	5.00
Standard deviation	Consistent	1.93
	Inconsistent	1.93
Kurtosis	Consistent	2.56
	Inconsistent	2.61
Skewness	Consistent	-0.10
	Inconsistent	-0.10

The consistent group had a mean score of 5.30, with a median of 5.0 and a standard deviation of 1.93. The inconsistent group had a mean score of 5.23, with a median of 5.0 and a standard deviation of 1.93. The analysis results indicate that both groups have similar descriptive statistics. The mean score for the consistent group was 5.30, with a median of 5.0 and a standard deviation of 1.93, while the inconsistent group had a mean score of 5.23, a median of 5.0, and a standard deviation of 1.93. The Wilcoxon rank-sum test (conducted due to lack of normal distribution of data) for independent groups showed that the difference between the mean scores of the consistent and inconsistent groups is not statistically significant ($W = 1278634$), ($p = 0.33$), which suggests no significant difference between the groups and, consequently, no confirmation of the hypothesis which stated that consistent face-name pairs will be viewed more favorably than inconsistent pairs even after exclusion of psychology students

Discussion

The studies revealed that the distributions of variables for both groups deviated from normality, as confirmed by the low p-values from the Shapiro-Wilk test. Both groups in both studies showed similar mean scores and standard deviations, as well as comparable skewness and kurtosis values. The lack of significant differences between the groups suggests that matching a name to a face did not have a significant impact on the likability ratings. It is worth noting that sound properties are likely a better predictor of the replication effect than speech properties (Passi & Arun, 2022). Our study did not employ sounds and was based solely on language. The phonology of a given country also appears to be important. For example, people who speak the Syuba language from Nepal did not show a preference for assigning shapes to sounds (Styles & Gawne, 2017). It is possible that Polish sounds are simply very similar in terms of “sharpness,” and even the most deviant ones we identified are not extreme enough to produce the effect (Styles & Gawne, 2017). A study comparing French and English names also found that orthography and font do not matter for the effect itself, but phonetics and auditory presentation (which we did not use in our study) best elicit the effect (Sidhu et al., 2016). It should also be mentioned that the original study used drawings, whereas we used AI-generated

photos to naturalize the obtained results. Of course, the small sample size in the pilot study may be not representative enough, and the fact that psychology students are required to pass a course that demands earning SONA credits can lead to the rejection of many responses due to failing the attention test, which may suggest a lack of attention during the study, potentially allowing some inattentive participants who happened to notice the attention check question to skew the results (Stawiany et al., 2024). In our study, we did not control for hearing problems or developmental disorders, and we have reason to believe that individuals who were deaf in childhood and received hearing aids later in life (Gold & Segal, 2019), as well as individuals on the autism spectrum (Shukla, 2016), may struggle with tasks related to the replicated effect. Given the lack of confirmation of such a widely described effect, after considering the above findings, it is recommended to conduct a preliminary phonetic study of the given names on a large group of people before conducting the main study, to ensure that the tools are adapted to the phonemes of the language and not just the letters, and to play sounds instead of showing written words. Additionally, it is important to ensure that participants are focused and to provide appropriate conditions for this, as well as to consider conducting autism screening tests and audiometry to exclude hearing problems.

Contributions: Contributed to conception and design: KS, KK, ZS; Contributed to acquisition of data: KS, KK; Contributed to analysis and interpretation of data: KS, KK; Drafted and/or revised the article: KS; Approved the submitted version for publication: KS, KK, ZS.

Acknowledgements: We would like to thank Dr. Angelika Olszewska from SWPS University for being our mentor and staying in touch with us even during the holidays, and for broadening our perception of statistics. We would also like to thank Dr. Kamil Izydorczak from SWPS University for helping us with the administrative process. Finally we really appreciate the help with the R language from Adrianna Gottfried who was the first person that explained us the R in the way we finally understand it. She did it better for us than some of the researchers.

Funding: We didn't receive any fundings from anyone.

Data Availability Statement: All of our work and our data, preregistration and R scripts are accessible by Open Science Framework under the name of "From "Bob/Kirk" to "Błażej/Piotrek" effect - replication of social bias effect towards people whose names match their faces". We also provide DOI of our project 10.17605/OSF.IO/87KRM.

Conflicts of Interest: Our team does not have any conflict of interests with anyone in publishing this article.

References

- Barton, D. N., & Halberstadt, J. (2018). A social Bouba/Kiki effect: A bias for people whose names match their faces. *Psychonomic Bulletin & Review*, 25, 1013-1020.
- Ćwiek, A., Fuchs, S., Draxler, C., Asu, E. L., Dediu, D., Hiovain, K., ... & Winter, B. (2022). The bouba/kiki effect is robust across cultures and writing systems. *Philosophical Transactions of the Royal Society B*, 377(1841), 20200390.
- Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39, 175-191.
- Farkas, L. G. (1994). *Anthropometry of the head and face*. Raven Pr.
- Gold, R., & Segal, O. (2020). The Bouba-Kiki Effect in persons with prelingual auditory deprivation. *Language Learning and Development*.
- Köhler, W. (1929). *Gestalt Psychology*. Liveright.
- Ministerstwo Cyfryzacji. (2020, Styczeń 29). Imiona nadawane dzieciom w Polsce w latach 2000-2019 - imię pierwsze. Otwarte Dane. https://dane.gov.pl/pl/dataset/219,imiona-nadawane-dzieciom-w-polsce?page=3&per_page=0&q=&sort=-data_date&model=resources
- Milfont, T. L., & Klein, R. A. (2018). Replication and reproducibility in cross-cultural psychology. *Journal of Cross-Cultural Psychology*, 49(5), 735-750.
- Passi, A., & Arun, S. P. (2022). The Bouba-Kiki effect is predicted by sound properties but not speech properties. *Attention, Perception, & Psychophysics*.

- Schmidt, F. L., & Oh, I. S. (2016). The crisis of confidence in research findings in psychology: Is lack of replication the real problem? Or is it something else? *Archives of scientific psychology*, 4(1), 32.
- Shukla, A. (2016). The Kiki-Bouba paradigm: where senses meet and greet. *Indian Journal of Mental Health*.
- Sidhu, D. M., Pexman, P. M., & Saint-Aubin, J. (2016). From the Bob/Kirk effect to the Benoit/Éric effect: Testing the mechanism of name sound symbolism in two languages. *Acta Psychologica*, 169, 88-99.
- Stawiany, K., Kozak, K., & Sokołowska, Z. (2024, August 24). From “Bob/Kirk” to “Błażej/Piotrek” effect - replication of social bias effect towards people whose names match their faces. <https://doi.org/10.17605/OSF.IO/87KRM>
- Styles, S. J., & Gawne, L. (2017). When does maluma/takete fail? Two key failures and a meta-analysis suggest that phonology and phonotactics matter. *i-Perception*.
- Usnadze, D. (1924). Ein experimenteller Beitrag zum Problem der psychologischen Grundlagen der Namengebung. *Psychologische Forschung*. Dostępne na: https://pure.mpg.de/rest/items/item_2403335/component/file_2403334/content
- Winter, B., Sóskuthy, M., Perlman, M., & Dingemanse, M. (2022). Trilled /r/ is associated with roughness, linking sound and touch across spoken languages. *Scientific Reports*, 12(1), 1035.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.