

# Realistic Image Primes for Experimental Research

Justin de Benedictis-Kessner\*

Tess Wise†

August 23, 2021

*Word count: 997*

## Abstract

Use of images as primes in political science experiments is widespread, but a standardized database of realistic diverse primes for this purpose has not been widely available to researchers. We develop an image database of faces of real people, which we call Realistic Image Primes for Experimental Research (RIPER). RIPER contains 249 images of individuals who are diverse along lines of gender, race, and occupational background. We standardize the images in RIPER by collecting ratings of their attractiveness, perceived race, and perceived income level. We make this diverse original database available as a useful tool for experimental social scientists.

Keywords: race, gender, class, identity, visual cues

---

We appreciate feedback on earlier versions of this manuscript from Daniel de Kadt and Krista Loose. This work was supported by the Institute for Quantitative Social Science (IQSS) at Harvard University and the Political Experiments Research Lab (PERL) at MIT. We received approval from the Committee on the Use of Humans as Experimental Subjects (COUHES) at MIT (protocol #1502006889) and the Committee on the Use of Human Subjects (CUHS) at Harvard.

\*Assistant Professor, John F. Kennedy School of Government, Harvard University, [jdbk@hks.harvard.edu](mailto:jdbk@hks.harvard.edu). Twitter: @jdbk.

†Assistant Professor, Department of Politics and International Affairs, Wake Forest University, [wiset@wfu.edu](mailto:wiset@wfu.edu). Twitter: @tesswise.

Researchers studying the effects of diversity on politics are confronted with a classic methodological tradeoff between internal and external validity. For example, in survey experiments which manipulate the race of a person in stimulus material, variation in attractiveness, age, or socio-economic status may all be accidentally manipulated along with the desired treatment of race, potentially due to a lack of “informational equivalence” (Dafoe, Zhang, and Caughey, 2018). Experimental subjects may respond to these other inadvertent manipulations (e.g., Lerman and Sadin 2016; McDermott 1998; Sen 2017; Stephens-Dougan 2016, though see Butler and Homola 2017). Race then becomes a compound treatment – which threatens the conclusions about experimental treatment effects being *caused* by race (Sen and Wasow, 2016). Examining the causal effects of diversity in experiments thus presents a particularly thorny methodological challenge of maintaining construct validity.

One way to circumvent this problem is for researchers to directly control some of these other factors alongside race when conducting experiments. In conjoint experiments, for instance, researchers often manipulate the race of a person about whom survey respondents are making a choice while also manipulating other target features such as occupation, religion, or political partisanship. However, these experiments usually must further abstract from reality by providing multiple explicit cues in an artificial format (e.g. Hainmueller, Hopkins, and Yamamoto, 2014). Research that improves on the problem of construct validity identified above may trade this for a loss of external validity.

Faced with this challenge alongside the substantive importance of studying diversity, researchers may seek to preserve external validity by using more realistic subtle visual cues (photos) rather than text cues about race. This choice presents several problems. For one, using visual rather than text cues may have treatment mode effects (Abrajano, Elmendorf, and Quinn, 2018). Moreover, using real photos ordinarily takes away some degree of experimental control – rendering experiments vulnerable to construct validity problems yet again. The logistical cost of collecting photos and then pre-testing and standardizing these photos to account for these problems may be (rightfully) daunting to experimental researchers.

We present one solution to this tradeoff with a diverse database of standardized image primes. We develop this database of faces, which we call Realistic Image Primes for Experimental Research (RIPER), from photos posted on public websites.<sup>1</sup> RIPER contains 249 images of individuals from three racial groups (Black, White, and Latino), and four different occupational groups (academics, graduate students, politicians, and maintenance workers). Included in RIPER alongside these photos are ratings of perceived attractiveness, perceived race, and perceived income level. Though other researchers have created databases of photo stimuli for experimental research in psychology, business, and other fields, these databases commonly rely on computer-generated faces (e.g. Matheson and McMullen, 2011), or are predominantly white (e.g. Lundqvist, Flykt, and Öhman, 1998), and usually feature images of people in identical outfits against sanitized backgrounds (e.g. Ma, Correll, and Wittenbrink, 2015; Tottenham et al., 2009). While these databases may be suitable if diversity along multiple dimensions is not a feature of interest, they are not as helpful for researchers studying the intersectional effects of diversity in a realistic manner related to politics. In particular, researchers that seek to use images of political candidates, policy beneficiaries, or voters in within-subjects designs, or using multiple images alongside one another that do not appear completely identical in outfit and background are left without many viable options. In contrast, the images in RIPER allow researchers to improve the realism of their experiments while controlling for potential confounders such as attractiveness and socioeconomic status that could weaken construct validity.

RIPER contains images of 249 real people taken from publicly available websites, which we standardize in resolution and size, cropped to show the individual only from the shoulders and above, and all of which depict a person against a plain background. Table 1 shows the number of photos within each demographic group in RIPER.

We had each image rated along several politically-relevant dimensions by workers on Amazon.com’s Mechanical Turk (MTurk). 1,507 unique MTurk workers reported their per-

---

<sup>1</sup>We describe the image collection process in Appendix A.

Table 1: Images by Profession, Gender, and Race

		Racial Group			Gender		
		Black	Latino	White	Female	Male	Total
Professional Group	Academic	13	11	47	25	46	71
	Grad Student	30	25	22	48	29	77
	Professional	10	9	28	25	22	47
	Working Class	11	15	28	25	29	54
Gender	Female	36	30	57			123
	Male	28	30	68			126
Total		64	60	125	123	126	249

ceptions of the attractiveness, race, and income levels of 10 faces selected randomly from our database.<sup>2</sup> We also used artificial intelligence from Microsoft Cognitive Services to machine-calculate several other photo characteristics, and hand-coded each image along several other dimensions ourselves, as described in Appendix D.

## Conclusion

The political effects of diversity are of critical importance for the discipline of political science, yet their study presents challenges for researchers who want to preserve both internal construct validity and external validity. We contribute a tool for researchers who wish to use realistic image primes while manipulating politically-relevant characteristics. Potential applications for this database are myriad: if researchers want to assess reactions to candidates, constituents, or policy beneficiaries of different races, these images may be a useful way to blend realistic information about race with experimental control. These images might also be used to complement existing experimental research that varies text information on race, or used to vary information about race alongside other features to assess its intersectional effects (e.g. Cassese and Barnes, 2019). We present results from two such applications, on jury sentencing decisions and candidate favorability, in Appendix E. If questions such as these are to be rigorously examined, then researchers should seek to use stimuli with conscious at-

<sup>2</sup>The full prompt used for the rating task is presented in Appendix B. Appendix C describes the sample of MTurk raters.

tention to both external and internal validity. The RIPER database facilitates investigation of diversity in ways that make conclusions more readily applicable to real-world phenomena.

## References

- Abrajano, Marisa A, Christopher S Elmendorf, and Kevin M Quinn. 2018. “Labels vs. Pictures: Treatment-Mode Effects in Experiments About Discrimination.” *Political Analysis* 26(1): 20–33.
- Butler, Daniel M, and Jonathan Homola. 2017. “An Empirical Justification for the Use of Racially Distinctive Names to Signal Race in Experiments.” *Political Analysis* 25(1): 122–130.
- Cassese, Erin C, and Tiffany D Barnes. 2019. “Intersectional Motherhood: Investigating Public Support for Child Care Subsidies.” *Politics, Groups, and Identities* 7(4): 775–793.
- Dafoe, Allan, Baobao Zhang, and Devin Caughey. 2018. “Information Equivalence in Survey Experiments.” *Political Analysis* 26(4): 399–416.
- Hainmueller, Jens, Daniel J Hopkins, and Teppei Yamamoto. 2014. “Causal Inference in Conjoint Analysis: Understanding Multidimensional Choices via Stated Preference Experiments.” *Political Analysis* 22(1): 1–30.
- Lerman, Amy E, and Meredith L Sadin. 2016. “Stereotyping or Projection? How White and Black Voters Estimate Black Candidates’ Ideology.” *Political Psychology* 37(2): 147–163.
- Lundqvist, Daniel, Anders Flykt, and Arne Öhman. 1998. “The Karolinska Directed Emotional Faces (KDEF).” CD ROM from Department of Clinical Neuroscience, Psychology section, Karolinska Institutet.
- Ma, Debbie S, Joshua Correll, and Bernd Wittenbrink. 2015. “The Chicago Face Database: A Free Stimulus Set of Faces and Norming Data.” *Behavior Research Methods* 47(4): 1122–1135.
- Matheson, Heath E, and Patricia A McMullen. 2011. “A Computer-Generated Face Database

- with Ratings on Realism, Masculinity, Race, and Stereotypy.” *Behavior Research Methods* 43(1): 224–228.
- McDermott, Monika L. 1998. “Race and Gender Cues in Low-Information Elections.” *Political Research Quarterly* 51(4): 895–918.
- Sen, Maya. 2017. “How Political Signals Affect Public Support for Judicial Nominations: Evidence from a Conjoint Experiment.” *Political Research Quarterly* 70(2): 374–393.
- Sen, Maya, and Omar Wasow. 2016. “Race as a Bundle of Sticks: Designs That Estimate Effects of Seemingly Immutable Characteristics.” *Annual Review of Political Science* 19.
- Stephens-Dougan, LaFleur. 2016. “Priming Racial Resentment without Stereotypic Cues.” *Journal of Politics* 78(3): 687–704.
- Tottenham, Nim, James W Tanaka, Andrew C Leon, Thomas McCarry, Marcella Nurse, Todd A Hare, David J Marcus, Alissa Westerlund, BJ Casey, and Charles Nelson. 2009. “The NimStim Set of Facial Expressions: Judgments from Untrained Research Participants.” *Psychiatry Research* 168(3): 242–249.