Educational Psychology Papers and Publications   Educational Psychology, Department of

2017

Ecological Invalidity of Existing Gaydar Research: In-Lab Accuracy Translates to Real-World Inaccuracy: Response to Rule, Johnson, & Freeman (2016)

William T.L. Cox
University of Wisconsin–Madison, will.cox@me.com

Patricia G. Devine
University of Wisconsin–Madison, pgdevine@wisc.edu

Alyssa A. Bischmann
University of Nebraska–Lincoln, abischmann2@unl.edu

Janet S. Hyde
University of Wisconsin–Madison, jshyde@wisc.edu

Follow this and additional works at: http://digitalcommons.unl.edu/edpsychpapers

Part of the Child Psychology Commons, Cognitive Psychology Commons, Developmental Psychology Commons, and the School Psychology Commons

Ecological Invalidity of Existing Gaydar Research: In-Lab Accuracy Translates to Real-World Inaccuracy: Response to Rule, Johnson, & Freeman (2016)

William T. L. Cox and Patricia G. Devine
Department of Psychology, University of Wisconsin–Madison

Alyssa A. Bischmann
Department of Educational Psychology, University of Nebraska–Lincoln

Janet S. Hyde
Departments of Psychology and Gender/Women's Studies, University of Wisconsin–Madison

Corresponding author — William T. L . Cox, University of Wisconsin–Madison, Department of Psychology,
1202 W. Johnson St., Madison, WI 53706. E-mail: will.cox@me.com

Abstract

In recent years, several empirical studies have claimed to provide evidence in support of the popular folk notion that people possess "gaydar" that enables them to accurately identify who is gay or lesbian (Rule, Johnson, & Freeman, 2016). This conclusion is limited to artificial lab settings, however, and when translated to real-world settings this work itself provides evidence that people's judgments about who is gay/lesbian are not pragmatically accurate. We also briefly review evidence related to the consequences of perpetuating the idea of gaydar (i.e., "the gaydar myth"). Although past claims about accurate orientation perception are misleading, the work that gave rise to those claims can nevertheless inform the literature in meaningful ways. We offer some recommendations for how the evidence in past "gaydar" research can be reappraised to inform our understanding of social perception and group similarities/differences.

One of the first idioms English-speaking children learn is that "You can't judge a book by its cover." This metaphorical phrase is meant to convey the principle that appearance is not a good and reliable way to make judgments about others. How people look, whether in regard to their attractiveness, their clothing, or the color of their skin, is but a very thin slice of a whole, complex human being, making it unlikely that any one attribute or set of attributes can accurately convey the complexity of another person. This idiom is likely so common because snap judgments, which often arise from stereotypes and implicit biases, are difficult to overcome (Cox, Abramson, Devine, & Hollon, 2012; Devine, 1989; Devine, Forscher, Austin, & Cox, 2012). A key reason for this difficulty is the existence of legitimizing myths—personal or cultural narratives that, one way or another, support reliance on processes such as stereotyping, snap judgments, or intuitions (e.g., Chen & Tyler, 2001; Glaser, 2005; Pettigrew, 1979; Quist & Resendez, 2002; Sidanius, Devereux, & Pratto, 1992; Uhlmann & Cohen, 2007). In a recent article (Cox, Devine, Bischmann, & Hyde, 2016), we identified how one such cultural narrative, "the gaydar myth," perpetuates stereotyping to infer sexual orientation.

The Gaydar Myth and Stereotyping to Infer Orientation

Because sexual orientation is not a visible group status, people often rely on stereotypic attributes commonly associated with gay men and lesbian women, such as fashion or career choice, to make snap judgments about who is gay or lesbian (Cox & Devine, 2015). Whereas stereotyping is often considered inappropriate (Cox & Devine, 2015; Devine, 1989; Devine & Elliot, 1995), giving this specific stereotyping process the alternate label of "gaydar" makes it seem acceptable. In other words, the gaydar myth *legitimizes* stereotyping to infer orientation. In a high-powered experiment (Cox et al., 2016, Study 5; achieved power for observed effect, $1 - \beta > 0.999$), we demonstrated that, compared to a control group, people who were led to

believe that "gaydar is real" relied more heavily on stereo-types to categorize men as gay. People who were told that gaydar is merely another term for stereotyping, however, stereotyped at much lower rates, even though they were neither discouraged from stereotyping nor told that the stereotypes were inaccurate. These results directly dem-onstrate that the folk notion of gaydar serves the function of a legitimizing myth: Rhetoric that authenticates gaydar increases stereotyping, but identifying gaydar *as* stereo-typing decreases reliance on stereotypes.

The people who rely most on stereotype-based gaydar are also more motivated to express prejudice (Forscher, Cox, Graetz, & Devine, 2015), and using stereotypes to pri-vately make a snap judgment that someone is gay grants "plausible deniability" to express anti-gay aggression (Cox & Devine, 2014). In an aggression study using real elec-tric shocks, a key subset of people took advantage of this plausible deniability as a smokescreen for their anti-gay prejudice, administering high levels of electric shocks to a man who was stereotypically implied to be gay (Cox & Devine, 2014). Although the cultural notion of gaydar may seem lighthearted, evidence indicates that the stereotyp-ing processes camouflaged under the guise of gaydar can have pernicious effects.

Inaccuracy of Gaydar

Whether or not stereotyping relates to other adverse con-sequences, it is possible for stereotypes to yield accurate snap judgments. In fact, as reviewed extensively by Rule and colleagues (2016) in reply to our recent article (Cox et al., 2016), several researchers claim that people's snap judgments about sexual orientation are largely accurate (e.g., Gaudio, 1994; Johnson, Gill, Reichman, & Tassinary, 2007; Rieger, Linsenmeier, Gygax, Garcia, & Bailey, 2010; Rule, Ambady, Adams, & Macrae, 2008). We argued that the conclusions of this past work were misleading and in-correct, based on (a) a methodological confound we ser-endipitously uncovered and (b) a fundamental flaw in the reasoning and design of the past work. As we review next, Rule, Johnson, and Freeman's (2016) reply falls short of addressing either of these issues.

Internal Validity Problems in Past Face-Based Gaydar Research

As reported in Cox et al. (2016), we found a quality con-found in two stimulus sets collected in an attempt to rep-licate and extend some of Rule et al.'s (2008) work on perceiving orientation from the face (i.e., "face-based gay-dar"). We discovered a natural confound, such that gay

men's and lesbian women's pictures were of higher quality than those of their straight counterparts. When we con-trolled for quality statistically or matched the stimuli on quality, the "gaydar" effects found in previous work disap-peared. Rule et al. (2016) did not present any data or argu-ments that disputed or tried to explain this nonreplication. In response to the possibility that their stimulus sets may contain the same quality confound, they collected data on the quality of their stimuli, finding that quality confounds existed in only five of their 13 stimulus sets. Several of the confounds in their stimulus sets were in the opposite di-rection of the confound we discovered, which they erro-neously offered as evidence against our claims. Rule and colleagues (2016) seem to misunderstand our original ar-ticle. The quality confounds, in our stimuli and in theirs, are a serious threat to internal validity, not a pattern that we theoretically predicted in a particular direction. Any ex-periment with built-in confounds is inherently undermined by those confounds, no matter their direction. With our two stimulus sets and Rule and colleagues' (2008) 13 sets, in total 47% of the stimulus sets contained quality con-founds, which does not inspire confidence about the in-ternal validity of this area of work.
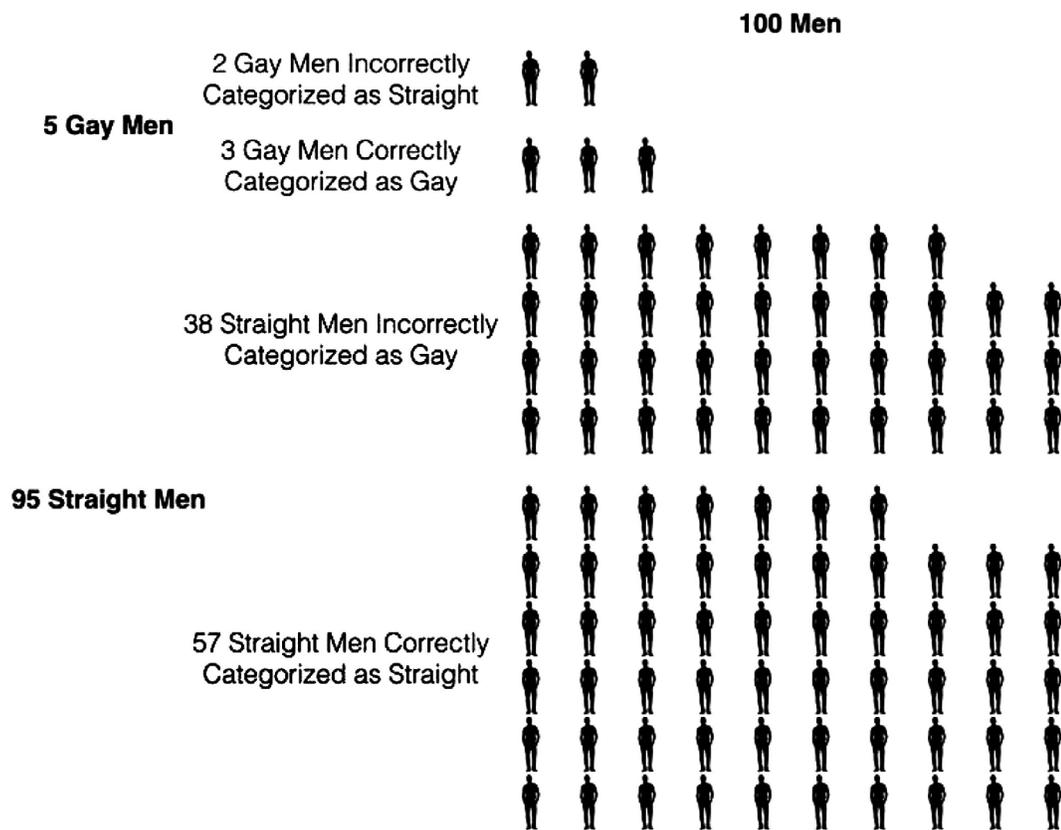
Quality is but one confound that may exist among stimulus sets. Because Rule and colleagues (2008) do not share their stimuli with outside research teams, other sci-entists cannot evaluate any other dimensions on which their stimuli may differ. Nevertheless, even if further stud-ies can address these internal validity threats and replicate past patterns of "accurate" orientation perception, doing so will not address the more fundamental flaw inherent in the reasoning underlying this area of work.

Evidence of Accuracy in the Lab Provides Evidence of Inaccuracy in the Real World

Rule et al. (2016) repeatedly state that our article ignored large swaths of evidence demonstrating the accuracy of gaydar. This claim, however, is incorrect. Although we did not specifically cite and review every paper they men-tioned, we concretely identified a mathematical and logi-cal error that is shared by all past gaydar research (see Cox et al., 2016, pp. 167–168; see also Plöderl, 2014). In so do-ing, we directly demonstrated that the patterns heralded as evidence of accurate gaydar in fact provide direct evi-dence that gaydar is highly inaccurate.

Gaydar studies largely follow the same basic formula, as follows: Participants are asked to make gay-or-straight judgments based on stimuli (e.g., pictures, video, sound clips) gathered from gay men and straight men.[1] Half of the stimuli come from straight men and half come from

---

1. We crafted our example with gay and straight men, for simplicity and to match the majority of past work reporting to demonstrate accurate gay-dar. The same mathematical principles and conclusions apply, however, to studies about categorization of lesbian and straight women, and to the rare studies that examine bisexual categorization.

100 Men

2 Gay Men Incorrectly
Categorized as Straight

5 Gay Men

3 Gay Men Correctly
Categorized as Gay

38 Straight Men Incorrectly
Categorized as Gay

95 Straight Men

57 Straight Men Correctly
Categorized as Straight

Figure 1. Laboratory evidence of gaydar accuracy provides evidence of real-world inaccuracy. Past gaydar research typically reports a 60% accuracy rate in lab studies using a false base rate in which 50% of the targets are gay men. This figure shows what 60% accuracy looks like when translated to a real-world situation in which only 5% of men are gay. Out of every 100 men, there will be 38 miscategorized straight men and three correctly categorized gay men. Based on all published evidence, using gaydar to identify someone as gay will be incorrect 93% of the time (38 / [38 + 3] = 92.7%).

gay men, and accuracy is assessed relative to 50% chance, which is the accuracy rate one would get if participants were purely guessing. In this work, people typically have 55% to 65% accuracy, which is significantly higher than 50% chance. Based on this evidence, then, do people have accurate gaydar? At first glance, the answer seems to be yes; people are correct more than they are incorrect. Further thought, however, reveals that this conclusion is erroneous.

The fundamental error in this work is its reliance on artificial base rates, in which 50% of the targets are gay. In the real world, the best estimates we have say that, at most, 5% of men identify as gay (Savin-Williams, 2006). This consideration of ecological validity drastically changes the interpretation of the 60% accuracy rate.[2] Figure 1 shows what a 60% accuracy rate looks like when translated to the real world, where only 5% of men are gay. An accuracy rate of 60% means that 40% of straight men are miscategorized as gay, and this 40% far outnumbers the 5% of men who are actually gay. Once real-world base rates of gay men are considered, all available evidence from past work indicates that when perceivers "use

their gaydar" to conclude that someone is gay, they will be wrong 93% of the time. Overall accuracy of 60% translates to 7% accuracy for identifying who is gay. This statistic is readily derived from the data of all past gaydar studies, and it provides a very clear answer: There is *no* pragmatic accuracy to gaydar.

Next Steps: Saving the Baby From the Bathwater

Our critique of past gaydar research is most precisely focused on the claim that "people can accurately perceive who is gay." Based on all available evidence, this conclusion is mathematically incorrect. An alternate phrasing of this claim is that "people can accurately perceive orientation," which, arguably, seems more justified, given that 60% of the straight men are categorized correctly. But even this claim is misleading when one notes that if people assumed that everyone was straight, they would have 95% accuracy. Any alleged process for "perceiving orientation" is meaningless if it does not yield pragmatic accuracy for identifying members of the numerical minority from among the majority. Furthermore, we have shown

2. This mathematical reasoning involves a basic application of Bayes's theorem, which is necessary for understanding probabilities and accuracy (Hooper, 2013; McGrayne, 2011).

3. This research objective brings with it many other potential difficulties to consider, most especially representative sampling of sexual minorities (Harry, 1986).

experimentally that this claim perpetuates stereotyping to infer orientation (Cox et al., 2016), a stereotyping process that has adverse consequences (Cox & Devine, 2014; Forscher et al., 2015).

Although the conclusion that people can accurately perceive orientation is incorrect, the research that has been used to support that conclusion can still inform the literature in meaningful ways. As we noted previously (Cox et al., 2016, p. 168), lab studies with artificial base rates can be useful for exploring intergroup similarities and differences between straight people and lesbian, gay, and bisexual (LGB) people.[3] For example, learning about what factors cause sexual orientation has been one of the biggest areas of study for sexuality researchers (DeLamater & Hyde, 1998). Insomuch as these past studies document meaningful group similarities and differences that can build or test theories about the etiology of orientation, they are highly valuable to sexuality science (e.g., Le-Vay, 2011; Rieger et al., 2010).

Even if they yield no pragmatic accuracy, understanding the cues people use to make snap judgments about orientation also has tremendous potential to inform our understanding of social perception, bias, and discrimination. If someone's facial structure, gait, or voice leads to a snap judgment or tacit inference about orientation, what consequences does that have? Will it affect hiring outcomes? Will it lead to social avoidance, anti-gay aggression, or other prejudice-related behaviors? Do these snap judgments lead to additional stereotypic inferences, for instance, about abilities, proclivities, and personality traits? Although we have shown that using stereotypes to infer sexual orientation leads to meaningful behavioral outcomes (Cox & Devine, 2014), we are not aware of any research demonstrating that inferences based on the face, voice, or body movement lead to discriminatory or prejudicial outcomes. Inferences based on these traits will not be accurate in the real world, but it is clear from this prolific body of work that people do use them to make snap judgments about orientation. The behavioral processes that follow these snap judgments are open for future exploration. Even if people cannot always accurately judge a book by its cover, there is much we can learn from the fact that people often make those judgments anyway.

ORCID — William T. L. Cox
http://orcid.org/0000-0001-5978-8195

References

Chen, E. S., & Tyler, T. R. (2001). Cloaking power: Legitimizing myths and the psychology of the advantaged. In A. Y. Lee-Chai & J. A. Bargh (Eds.), *The use and abuse of power: Multiple perspectives on the causes of corruption* (pp. 241–261). New York, NY: Psychology Press.

Cox, W. T. L., Abramson, L. Y., Devine, P. G., & Hollon, S. D. (2012). Stereotypes, prejudice, and depression: The integrated perspective. *Perspectives on Psychological Science*, *7*, 427–449. doi:10.1177/1745691612455204

Cox, W. T. L., & Devine, P. G. (2014). Stereotyping to infer group membership creates plausible deniability for prejudice-based aggression. *Psychological Science*, *25*, 340–348. doi:10.1177/0956797613501171

Cox, W. T. L., & Devine, P. G. (2015). Stereotypes possess heterogeneous directionality: A theoretical and empirical exploration of stereotype structure and content. *PLoS ONE*, *10*, e0122292. doi:10.1371/journal.pone.0122292

Cox, W. T. L., Devine, P. G., Bischmann, A. A., & Hyde, J. S. (2016). Inferences about sexual orientation: The roles of stereotypes, faces, and the gaydar myth. *Journal of Sex Research*, *53*, 157–171. doi:10.1080/00224499.2015.1015714

DeLamater, J. D., & Hyde, J. S. (1998). Essentialism vs. social constructionism in the study of human sexuality. *Journal of Sex Research*, *35*, 10–18. doi:10.1080/00224499809551913

Devine, P. G. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology*, *56*, 5–18. doi:10.1037/0022-3514.56.1.5

Devine, P. G., & Elliot, A. J. (1995). Are racial stereotypes really fading? The Princeton trilogy revisited. *Personality and Social Psychology Bulletin*, *21*, 1139–1150. doi:10.1177/01461672952111002

Devine, P. G., Forscher, P. S., Austin, A. T., & Cox, W. T. L. (2012). Long-term reduction in implicit race bias: A prejudice habit-breaking intervention. *Journal of Experimental Social Psychology*, *48*, 1267–1278. doi:10.1016/j.jesp.2012.06.003

Forscher, P. S., Cox, W. T. L., Graetz, N., & Devine, P. G. (2015). The motivation to express prejudice. *Journal of Personality and Social Psychology*, *109*, 791–812. doi:10.1037/pspi0000030

Gaudio, R. (1994). Sounding gay: Pitch properties in the speech of gay and straight men. *American Speech*, *69*, 30–57. doi:10.2307/455948

Glaser, J. (2005). Intergroup bias and inequity: Legitimizing beliefs and policy attitudes. *Social Justice Research*, *18*, 257–282. doi:10.1007/s11211-005-6825-1

Harry, J. (1986). Sampling gay men. *Journal of Sex Research*, *22*, 21–34. doi:10.1080/00224498609551287

Hooper, M. (2013). Richard Price, Bayes' theorem, and God. *Significance*, *10*, 36–39. doi:10.1111/j.1740-9713.2013.00638.x

Johnson, K. L., Gill, S., Reichman, V., & Tassinary, L. G. (2007). Swagger, sway, and sexuality: Judging sexual orientation from body motion and morphology. *Journal of Personality and Social Psychology*, *93*, 321–334. doi:10.1037/0022-3514.93.3.321

LeVay, S. (2011). *Gay, straight, and the reason why: The science of sexual orientation*. New York, NY: Oxford University Press.

McGrayne, S. B. (2011). *The theory that would not die: How Bayes' rule cracked the enigma code, hunted down Russian submarines, and emerged triumphant from two centuries of controversy*. New Haven, CT: Yale University Press.

Pettigrew, T. F. (1979). The ultimate attribution error: Extending Allport's cognitive analysis of prejudice. *Personality and Social Psychology Bulletin*, *5*, 461–476. doi:10.1177/014616727900500407

Plöderl, M. (2014). Bayesian advice for gaydar-based picking up: Commentary on Lyons, Lynch, Brewer, and Bruno (2013). *Archives of Sexual Behavior*, *43*, 7–9. doi:10.1007/s10508-013-0178-x

Quist, R. M., & Resendez, M. G. (2002). Social dominance threat: Examining social dominance theory's explanation of prejudice as legitimizing myths. *Basic and Applied Social Psychology*, *24*, 287–293. doi:10.1207/S15324834BASP2404_4

Rieger, G., Linsenmeier, J. A. W., Gygax, L., Garcia, S., & Bailey, J. M. (2010). Dissecting "gaydar": Accuracy and the role of masculinity–femininity. *Archives of Sexual Behavior*, *39*, 124–140. doi:10.1007/ s10508-008-9405-2

Rule, N. O., Ambady, N., Adams, R. B. Jr., & Macrae, C. N. (2008). Accuracy and awareness in the perception and categorization of male sexual orientation. *Journal of Personality and Social Psychology*, *95*, 1019–1028. doi:10.1037/a0013194

Rule, N. O., Johnson, K. L., & Freeman, J. B. (2016). Evidence for the absence of stimulus quality differences in tests of the accuracy of sexual orientation judgments: A reply to Cox, Devine, Bischmann, and Hyde (2016). *Journal of Sex Research*. Advance online publication. doi:10.1080/00224499.2016.1205547

Savin-Williams, R. C. (2006). Who's gay? Does it matter? *Current Directions in Psychological Science*, *15*, 40–44. doi:10.1111/cdir.2006.15.issue-1

Sidanius, J., Devereux, E., & Pratto, F. (1992). A comparison of symbolic racism theory and social dominance theory as explanations for racial policy attitudes. *Journal of Social Psychology*, *132*, 377–395. doi:10.1080/00224545.1992.9924713

Uhlmann, E. L., & Cohen, G. L. (2007). "I think it, therefore it's true": Effects of self-perceived objectivity on hiring discrimination. *Organizational Behavior and Human Decision Processes*, *104*, 207–223. doi:10.1016/j.obhdp.2007.07.001