Gaydar and the fallacy of decontextualized measurement*

Andrew Gelman† Greggor Mattson‡ Daniel Simpson§

17 Apr 2018

Abstract

Recent media coverage of studies about "gaydar," the supposed ability to detect another's sexual orientation through visual cues, reveal problems in which the ideals of scientific precision strip the context from intrinsically social phenomena. This fallacy of objective measurement, as we term it, leads to nonsensical claims based on the predictive accuracy of statistical significance. We interrogate these gaydar studies' assumption that there is some sort of pure biological measure of perception of sexual orientation. Instead, we argue that the concept of gaydar inherently exists within a social context and that this should be recognized when studying it. We use this case as an example of a more general concern about illusory precision in the measurement of social phenomena, and suggest statistical strategies to address common problems.

gaydar, *n.*

Pronunciation: Brit. `/'geida:/` , U.S. `/'gei,dar/`
Etymology: Blend of gay adj. and radar n.
slang.

An ability, attributed esp. to homosexual people and likened humorously to radar, to identify a (fellow) homosexual person by intuition or by interpreting subtle signals conveyed by appearance or behaviour.

— Oxford English Dictionary (2003)

"Gaydar" colloquially refers to the ability to accurately glean others' sexual orientation from mere observation. But does gaydar really exist? If so, how does it work?

Our research, published recently in the peer-reviewed journal PLoS ONE, shows that gaydar is indeed real and that its accuracy is driven by sensitivity to individual facial features as well as the spatial relationships among facial features.

— Tabak and Zayas (2012b)

1. Introduction

The science of detecting sexual orientation has experienced something of a renaissance, attracting researchers whose studies garner broad news coverage. The portmanteau of "gay" and "radar" first emerged in print among gay and lesbian comedians in the mid 1990s (for example, DiLallo and Krumholtz, 1994), becoming the name of a popular international dating website in 1999 and debuting on U.S. television through the comedies Will & Grace and Futurama (both in 1999) and Queer Eye for the Straight Guy (in 2003). Psychologists soon tested for the ability (Shelp, 2003), such as when a Harvard undergraduate's 2005 senior thesis garnered coverage in *Psychology Today* with the announcement, "It's true: Some people really do have 'gaydar'" (Lawson, 2005).

*To appear in *Sociological Science*. We thank Michal Kosinski and Martin Plöderl for helpful comments.

†Department of Statistics and Department of Political Science, Columbia University, New York.
‡Department of Sociology and Program in Gender, Sexuality, and Feminist Studies, Oberlin College, Ohio.
§Department of Statistical Sciences, University of Toronto, Canada.

Subsequent studies garnered international media attention, including "Advances in AI are used to spot signs of sexuality" (*Economist*, 2017), and "New study finds that your 'gaydar' is terrible" (Feltman, 2015).

One problem common to all of these studies, and to the breathless media coverage that followed them (see, for example, Schramm, 2018), was their overestimation of their import outside laboratory or modeled conditions. Stripping away all social context from an inherently social feature consistently produces results that imply the existence of an essential homosexual nature that can be detected, and that algorithms can be trained to perform more accurately than humans. We address these conceptual problems below after detailing the errors of statistical extrapolation that underpin them.

2. Low population frequency makes gaydar unreliable

We first review some mathematics that reveals the limitations of any attempt to identify membership in a group that forms only a small subset of the general population.

Let p be the proportion of adults in the United States who identify as gay. For simplicity, let us assume that $p = 4\%$ and that everybody is either gay or straight. (It would not be difficult to extend the model to allow for intermediate or other characterizations.) Now suppose that you have a gaydar which emits a continuous signal when you study a person, and the reflection of that signal contains a signature that informs you that the person is gay. Or, to put it more prosaically, you process some number of attributes conveyed by a person which enable a probabilistic classification, assuming you have been accurately trained and can correctly adjust for changes in base rate.

To simplify, suppose you are required to compress your continuous gaydar measurement into a binary go/no-go decision using some threshold which will trade off between false positives and negatives. Let α be the probability that a gay person is correctly classified as gay, and let β be the probability that a straight person is correctly classified as straight; thus with perfect gaydar, $\alpha = \beta = 1$.

Suppose you classify someone as gay; then the probability that he or she actually is gay is $\Pr(\text{gay given they have been classified as gay}) = \frac{p\alpha}{p\alpha+(1-p)(1-\beta)}$. It is a well known result from conditional probability that when p is low, the misclassification rate is high. For example, if $p = 0.04$ and $\alpha = \beta = 0.9$ (which would be an impressive accuracy in many settings), then $\Pr(\text{gay} \,|\, \text{classified as gay}) = \frac{0.04 \cdot 0.9}{0.04 \cdot 0.9 + 0.96 \cdot 0.1} = 0.27$; thus, you would actually be wrong nearly three-quarters of the time. Under this system you would be classifying 13.2% of people as gay—plausible given that people generally overestimate the proportions of rare events (see Hemenway, 1997) and groups, including immigrants, ethnic minorities, and gays (see Sides and Citrin, 2007, Newport, 2015, and Srivastava, 2011).

From these considerations, we can also work out how accurate a classifier needs to be on straight and gay populations to achieve any specified classification accuracy of gay people when applied to a population where the proportion of homosexuals is p. If we want the chance of someone who is classified as gay actually being gay to be greater than 50%, then we need $\beta > 1 - \frac{p\alpha}{1-p}$ to hold. When p is small, this means that β, the probability of classifying a straight person correctly, will have to be very close to 1. For example, if 4% of the population is gay, then even if we could identify every single gay person perfectly, we would still need to classify around 96% of straight people correctly to achieve the required accuracy. The intuition under this result is that because there are so many more straight people than there are gay people, there is an enormous penalty if they are incorrectly classified. This error underlies two recent gaydar studies, which also suffered from conceptualizing homosexuality, and the signals sent about it, as asocial phenomena. The point here is not that

partial classification is impossible but rather that it will be challenging to have a high probability of correct prediction in any particular case.

3. Two gaydar experiments

In a study that received extensive media attention, Tabak and Zayas (2012a) assessed the abilities of 24 college-student volunteers at identifying the sexual orientations of 400 self-identified gay or straight people based on photographs of faces that excluded images of individuals with eyewear, jewelry, scars, or other "facial alterations."[1] Half the targets in the study were gay and half were straight, and the students correctly identified sexual orientation 60% of the time, which was statistically significantly better than the 50% that would be expected by pure guessing.

Tabak and Zayas write that their research "was the first attempt to determine the roles that featural and configural face processing play in snap judgments of sexual orientation from faces," and it indeed seems to provide a clue about such visual manifestations. But we disagree with their claim that they have shown that "configural face processing significantly contributes to perception of sexual orientation."

To understand our disagreement, consider several aspects of gaydar as we understand it, and which are consistent with the dictionary definition given at the start of this paper. Gaydar occurs in social contexts with information including voice, dress, posture, and even topics of conversation or the places and contexts in which they occur; the Tabak and Zayas study removes all such cues. Gaydar is relevant in settings where gays are a small fraction of the population; in contrast, gay people represented 50% of the photos in the experiment under discussion. Indeed, Tabak and Zayas act as if they have removed all social cues, leaving behind nothing but the asocial and objective face, as if photographs are taken in laboratory conditions or that smiles or eye expressions are not themselves "facial alterations." Gaydar under these laboratory conditions has been transmuted from the in-group task of identifying fellow members of a rare subgroup in social interactions, to a mechanical binary classification task of asocial faces conceived as neutral.

More recently, Wang and Kosinski (2018) performed a similar exercise, this time using a machine learning algorithm to identify faces as gay or straight using 35,326 images scraped from an unnamed dating website. Once again, approximately half of these images are of people who were using the website to find members of the same sex. Their algorithm was able to classify sexual orientation correctly in 70–80% of a subset of this dataset that was held aside while building the classifier. The classifier also worked, although less well, on a set of around 900 faces of white men from Facebook who both identified as looking for a male partner and liked at least two Facebook pages such as "I love being gay" and "Gay and fabulous," and in other experiments they also compared to human judges and to computer classification using landmarks relating to facial morphology.

Here we focus on the classification of the images from the dating website. It can be interesting to see what happens to show up in the data (lesbians wear baseball caps! gay men have less facial hair!), but we question their extrapolation from their experimental conditions into speculations regarding inherent associations of homosexuality, for example:

> "it is unclear whether gay men were less likely to wear a beard because of nature (sparser facial hair) or nurture (fashion). If it is, in fact, fashion (nurture), to what extent is

[1] "To minimize the prospect that non-face cues would influence judgments, photographs of men or women with facial alterations or adornments (e.g., scars, eyewear, facial hair, makeup, non-earlobe piercings, etc.) were not included as experimental targets. To maximize consistency across faces, only photographs of White-appearing individuals who self-identified ages of 18–29 were included. Using Adobe Photoshop CS3 Extended, research assistants then removed the hair and ears from each head and converted the images to grayscale (8-bit bitmap format) to create a standardized 'face' stimulus."

such a norm driven by the tendency of gay men to have sparser facial hair (nature)? Alternatively, could sparser facial hair (nature) stem from potential differences in diet, lifestyle, or environment (nurture)?" (p. 254).

Similarly, the authors suggest that the correlation between facial brightness and probability of being gay could be evidence that straight men have higher levels of testosterone. An alternative reading of that result is that gay men are more likely to postprocess their dating profile pictures using a variety of readily available tools and filters, part of the impression management strategies that are rife in online profile photographs (Toma and Hancock, 2010, Zytko, Grandhi, and Jones, 2014). As Cohen (2017) and Mattson (2017) note, the speculation that Wang and Kosinski engage in is essentially disconnected from the data analysis that is being used as its justification.

Wang and Kosinski report that their goal was to "advance our understanding of the origins of sexual orientation and the limits of human perception ...," but their data provide no evidence regarding the former. As for the latter, a challenge is that the people in the photographs may well be intentionally using their facial expressions to communicate sexual orientation, given that people classified in one of the studies were selected based on Facebook "likes" on such pages as "Gay Times Magazine" and "Manhunt" as evidence for actual sexual orientation.

Neither Wang and Kosinski nor Tabak and Zayas produced results that suggest that more than half of the people flagged by their respective laboratory gaydars would be gay if these gaydars had been applied to the general population, and that is consistent with results from Olivola and Todorov (2010) on the error rates of sexual orientation judgments applied in realistic social settings. In stating these limitations, we do not intend to reject the empirical results of Tabak and Zayas (2012a) and Wang and Kosinski (2018). Their classification results are interesting and may well point to new insights regarding variation in facial expression.

4. Sampling and social context

Both the studies under discussion here measure the perception of sexual orientation in a context-free way. Decontextualization—bringing a phenomenon "into the lab" for careful study—is a characteristic step of scientific measurement, but it can cause problems in fields such as ecology and social science, where context is all. Reductionism—breaking a complex phenomenon into simpler parts to enable understanding—is a necessary part of the scientific enterprise, but bracketing the social for an inherently social phenomenon causes its evaporation, not its reduction. In particular, we have three concerns with these laboratory studies of gaydar.

The easiest concern to state is representativeness. A low frequency of facial hair among openly gay men who post to a particular dating website to find other gay men, or a finding that "sexual orientation is inferred more easily from women's vs. men's faces," may well be telling us more about the samples than about the general population they are presumed to represent. These are what Magnet (2011) calls the "demographic failures" baked into biometric technologies by reductive and biased samples; after an exhaustive review of their failures, she concludes that "human bodies are not biometrifiable." Given that no census or representative sample exists of images of gay people (or, for that matter, straight people), any statistical analysis will always have to deal with the extrapolation problem, and we recommend using some sort of multilevel model that explicitly allows for variation among and within different subgroups of each population. After all, there is great variation among even heterosexuals, some of whose "hybrid" expressions—such as metrosexual men or working-class Midwestern women—appear to others to be homosexual when they are not (Hall, 2015, Bridges and Pascoe, 2014, Kayzak, 2012).

Our second concern is the way in which gaydar, which was originally framed as an aspect of

communication *within* the gay community, has been redefined in the news media as a skill that can be deployed by the general (thus, mostly straight) population. One can distinguish between "active" gaydar (in which members of a subgroup are sending coded messages to each other—what Shelp, 2003, calls adaptive gaydar) and "passive" gaydar, in which outsiders catch some of these signals even though they are not the intended recipients. *In either case*, we suspect that many if not most of the distinctive and noticeable characteristics of the subpopulation are the result of active choices by members of that group, not (as assumed in the two papers under discussion) essential attributes derived from the false binary of "nature" or "nurture."

By taking gaydar into the lab, these research teams have taken the creative adaptation of an oppressed community of atomized members and turned gaydar into an essentialist story of "gender atypicality," a topic that is related to, but distinctly different from, sexual orientation (see, for example, Valdes, 2005, Fausto-Sterling, 2000, and Newton, 1984). This new story has moved gay people from the protagonists of the story to research objects in which there is the potential for construction of just-so stories based on of gender stereotypes. Again, a certain amount of reduction and objectification is necessary in scientific research—but researchers must be aware of what is lost in these steps in generalization to real-world settings.

Our third concern is the use of gender stereotypes in the deduction of homosexuality. Researchers have long distinguished between same-sex identity, same-sex desires, and same-sex behavior, distinct phenomena that "are imperfectly correlated and inconsistently predictive of each other" (Savin-Williams, 2006; see also Laumann et al., 1994). The relationship between gender identity and gender expression is similarly fraught in relation to sexual orientation. Researchers who attend to such distinctions have not found a reliable gaydar among research subjects, such as the finding that "the stereotypic association of feminine looking men as homosexual may confound judgments of sexual orientation" (Valentova, 2014). Other research is definitive, finding that "stereotypes casting gays and lesbians as gender 'inverts,' in cultural circulation for a century and a half, lead perceivers to use gendered facial cues to infer sexual orientation" (Freeman et al., 2010; see also Rieger, 2010). In other words: people can readily detect gender atypicality, but its relationship to homosexuality—and heterosexuality—is unclear. Given the consistent evidence that same-sex identity is organized differently among men and women, and also that bisexuality may be that most common expression of homosexuality among women (Bailey et al., 2016), it is problematic on gender grounds to reduce sexuality orientation to a binary variable, a reduction that is understandable for the simplicity of modeling but should be carefully noted when discussing these studies.

Gaydar is itself a mode of communication that is contingent on social structures and expectations. Colzato et al. (2010) find that subjects who are told that gaydar exists will have it; they also have found that homosexual subjects attend more closely to patterns and details than heterosexuals, which they presume "increases the likelihood to detect perceptual cues indicative of orientation, which again facilitates finding like-minded, social peers, and potential friends and sex mates." Their research is consistent with the idea of gaydar as a form of communication, and not some intrinsic radiation from flaming homosexuals.

5. The fallacy of decontextualized measurement

The reporting and interpretation of the gaydar experiments suffered from three problems discussed below which are common in social science research. We can identify all these problems with what might be called *the fallacy of decontextualized measurement*, the idea that science proceeds by crisp distinctions, modeled after asocial phenomena such as unambiguous medical diagnoses (the presence or absence of streptococcus, or the color change of a litmus paper). Seeking an on-off

decision, normalizing a base rate to 50%, and, most problematically, stripping a phenomenon of its social context: all these give the feel of scientific objectivity while creating serious problems for generalizing findings to the world outside the lab or algorithm.

The first problem we noticed in these studies is the reduction of an (implied) continuous scale to a binary choice. Some people appear clearly gay, others emit some gay signals, while others appear completely straight. Gaydar is on a sliding scale and depends on context: again, the traditional goal is to identify gay people who might be signaling their sexual identities to the in-group while staying hidden from the general population, which is quite a bit different from a study such as Wang and Kosinski (2018), which used participants on a dating site who both self-identified as gay and want other people to know this. Some of the problems can be reduced by at least adding a third category to the classification that accounts for cases where the classifier (be it human or machine) is unsure. A guiding principle when trying to split a population into two groups is that you always need a third group to account for those individuals that are within a margin of error (Gelman and Park, 2009). Applying this "rule of thirds" would not fix the problem of reducing a continuous scale to a binary choice, but it does allow the researchers to better assess and communicate the uncertainty in their method.

Second, these laboratory experiments have a much different base rate than in real-world settings, a point also made by Plöderl (2014) and Cox et al. (2017). It is well known that judgments of uncertainty are contingent on base rate (Kahneman and Tversky, 1974), and this is particularly relevant for a concept such as gaydar which arises in a setting in which the challenge is identifying a small minority in a large population. This effect is obvious in the Tabak and Zayas study—where orientation was assessed by humans—and it could be partly mitigated by repeating the experiment with only 4% of the photographs portraying a homosexual. It could also be instructive to repeat the experiment using a set of photographs that are only of heterosexuals: if these studies are purely classificatory exercises, then the ability to spot heterosexuals would be as interesting as any other algorithm.

Adapting the Wang and Kosinki experiment to a low base rate presents a slightly different prospect. There is nothing wrong with using a 50/50 sample in the first step of their procedure, which extracts around 4000 facial features that are used to distinguish between the two cases. As these researchers note, these features can be converted into a classifier by adjusting the base rate. While these studies may have not intended to study categorization in real life but rather general perceptual capabilities (e.g., Bruno, Lyons, and Brewer, 2014), such careful discussion is often missing in study framings and discussions by researchers in the news media.

Third, the researchers took a rich real-world phenomenon and abstracted it so much that they removed all its social content (or believe they have: eyes are maintained, as findings that lesbians wear less eye makeup than straight women attest). Gaydar has traditionally existed within a particular social context—a world in which gays are an invisible minority, existing in plain sight and seeking to be inconspicuous to the general population while communicating with others of their subgroup using various coordination practices (see, e.g., Minnelli et al., 1989). Face shapes may tell us something about gender atypicality, and we see value to studies of human and machine classifications of faces, but interpretations of these empirical findings will be made in the context of social norms and the ways humans interact.

6. Discussion

In recent years psychologists have studied correlates of sexual orientation in a variety of ways that have attracted various media coverage (see, for example, France, 2007, and Saletan, 2011). The place of homosexuality in our culture has changed dramatically in recent decades, to the extent that

concepts such as gaydar are changing their meaning and probably also their practice. This suggests that the signaling of the past is different from that of today: compare the cagey flamboyance of male figure skater Johnny Weir in the 2008 Winter Olympics to the unabashed "faggy magic" of Adam Rippon in the games one decade later (Moskowitz, 2018).

As noted above, we are not claiming that experiments such as those of Tabak and Zayas (2012) and Wang and Kosinski (2018) are useless. If various aspects of the shapes of faces are (weakly) correlated with sexual orientation, even through the intervening variable of gender atypicality; and if untrained volunteers or computer programs can classify such patterns with high accuracy, this is possibly interesting. Some insight might be gained by performing a set of studies comparing other groups, each time using people or computer programs to classify people chosen from two different samples, for example college graduates and non-college graduates, or English people and French people, or driver's license photos in state X and driver's license photos in state Y, or students from college A and students from college B, or baseball players and football players, or people on straight dating site U and people on straight dating site V, or whatever. From a statistical standpoint, it could be useful to think of gay vs. straight as one of many possible classifications, and researchers could then explore the ways in which classification accuracy varies by group division and data source. Reductionism is a characteristic and useful tool of science when used to understand a complex phenomenon by breaking it down into simple parts. In these cases, however, too much has been stripped away from social reality to make any general conclusions about differences between gay and straight people.

The word "gaydar" never appeared in either of the two papers under discussion (except in some of their references), but the expression was all over the place in media reports as well as in critical discussions such as this one. This illustrates one of the challenges of research in this area, that scientific research does not take place in a vacuum. It is not paradoxical that two papers that never use the word "gaydar" can be all about gaydar—if this is how they are received. We thus emphasize that we are exploring these research papers and their scientific and popular receptions all at the same time. If popular (mis)conceptions of gaydar are warping the reception of serious scientific studies, this is something researchers must confront.

The point of this article is not to pick on a small area of psychology research that happened to catch the fancy of the press, or even to criticize larger trends of sensationalism in science and the news media. Rather, we seek to draw attention to the general problem, all to frequent in this era of genetics, machine learning algorithms, and MRI studies, in which the ideals of scientific precision end up stripping all context from a social phenomenon, leading to nonsensical claims based on predictive accuracy or statistical significance. Statistical and machine learning methods can be very useful in social science, but they are hard to use well and can be easily misinterpreted. These algorithms are only able to do one thing: they interpolate the training data. It is difficult to bracket social stereotypes which often form the taken-for-granted common sense, leading to Ceglowski's (2016) caution that "machine learning is like money laundering for bias" when artificial worlds are substituted for the social world. Indeed, many of our critiques have been leveled by psychologists and AI researchers at their colleagues (Fasoli and Hegarty, 2017, Cox et al., 2016, 2017). Restoring context may come from research partnerships, perhaps by having studies reviewed by experts in the social phenomena and not merely the methods, or even coauthoring with them. A social interaction cannot always be measured in a test tube or even in a psych lab. In the case of the research under discussion here, the steps taken ostensibly to ensure objectivity obliterated much of the objects of research.

References

Bailey, J. M., Vasey, P. L., Diamond, L. M., Breedlove, S. M., Vilain, E., and Epprecht, M. (2016). Sexual orientation, controversy, and science. *Psychological Science in the Public Interest* **17** (2), 45–101.

Bridges, T., and Pascoe, C. J. (2014). Hybrid masculinities: New directions in the sociology of men and masculinities. *Sociology Compass* **8**, 246–258.

Bruno, D., Lyons, M., and Brewer, G. (2014). Response to Bayesian advice for gaydar-based picking up: Commentary on Lyons, Lynch, Brewer, and Bruno (2013) by Plöderl. *Archives of Sexual Behavior* **43**, 11–12.

Ceglowski, M. (2016). The moral economy of tech. Society for the Advancement of Socio-Economics Annual Meeting, Berkeley, Calif., 26 Jun. `http://idlewords.com/talks/sase_panel.htm`

Cohen, P. (2017). On artificially intelligent gaydar. Family Inequality blog, 11 Sep. `https://familyinequality.wordpress.com/2017/09/11/on-artificially-intelligent-gaydar/`

Colzato, L. S., van Hooidonk, L., van den Wildenberg, W. P. M., Harinck, F., and Hommel, B. (2010). Sexual orientation biases attentional control: A possible gaydar mechanism. *Frontiers in Psychology* **1**, 13.

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

Cox, W. T. L., Devine, P., Bischmann, A., and Hyde, J. S. (2017). Ecological invalidity of existing gaydar research: In-lab accuracy translates to real-world inaccuracy: Response to Rule, Johnson, and Freeman (2016). *Journal of Sex Research* **54**, 820–824.

DiLallo, K., and Krumholtz, J. (1994). *The Unofficial Gay Manual: Living the Lifestyle (Or at Least Appearing to)*. New York: Doubleday.

Economist. (2017). Advances in AI are used to spot signs of sexuality, 9 Sep.

Fasoli, F., and Hegarty, P. (2017). Straight talk about gaydar: How do individuals guess others' sexual orientation? *The Inquisitive Mind* **7**, 34.

Fausto-Sterling, A. (2000). *Sexing the Body: Gender Politics and the Construction of Sexuality*. New York: Basic Books.

Feltman, R. (2015). New study finds that your 'gaydar' is terrible. *Washington Post*, 9 Sept.

France, D. (2007). The science of gaydar. *New York*, 17 Jun. `http://nymag.com/news/features/33520/`

Freeman, J. B., Johnson, K. L., Ambady, N., and Rule, N. O. (2010). Sexual orientation perception involves gendered facial cues. *Personality and Social Psychology Bulletin* **36**, 1318–1331.

Gelman, A., and Park, D. (2009). Splitting a predictor at the upper quarter or third and the lower quarter or third. *American Statistician* **63**, 1–8.

Hall, M. (2015). *Metrosexual Masculinities*. New York: Palgrave Macmillan.

Hemenway, D. (1997). The myth of millions of annual self-defense gun uses: A case study of survey overestimates of rare events. *Chance* **10** (3), 6–10.

Laumann, E. O., Gagnon, J. H., Michael, R. T., and Michaels, S. (1994). *The Social Organization of Sexuality: Sexual Practices in the United States*. University of Chicago Press.

Lawson, W. (2005). Queer eyes: Blips on the gaydar. *Psychology Today*, 1 Nov.

Kahneman, D., and Tversky, A. (1974). Judgment under uncertainty: Heuristics and biases. *Science* **185**, 1124–1131.

Magnet, S. A. (2011). *When Biometrics Fail: Gender, Race, and the Technology of Identity.* Duke University Press.

Mattson, G. (2017). Artificial intelligence discovers gayface. Sigh. Greggor Mattson blog, 9 Sep. https://greggormattson.com/2017/09/09/artificial-intelligence-discovers-gayface/

Minnelli, L., Tennant, N., Lowe, C., and Mendelsohn, J. (1989). *Results.* New York: Epic Records.

Moskowitz, P. (2018). The faggy magic of Adam Rippon. *Splinter*, 16 Feb.

Newport, F. (2015). Americans greatly overestimate percent gay, lesbian in U.S. *Gallup News*, 21 May. http://news.gallup.com/poll/183383/americans-greatly-overestimate-percent-gay-lesbian.aspx

Newton, E. (1984). The mythic mannish lesbian: Radclyffe Hall and the New Woman. *Signs* **9**, 557-575.

Olivola, C. Y., and Todorov, A. (2010). Fooled by first impressions? Reexamining the diagnostic value of appearance-based inferences. *Journal of Experimental Social Psychology* **46**, 315–324.

Oxford English Dictionary, third edition (2003). Definition of "gaydar." Oxford University Press. http://www.oed.com/viewdictionaryentry/Entry/247143

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.

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

Saletan, W. (2011). Read my lisp. *Slate*, 15 Jul. http://www.slate.com/articles/health_and_science/human_nature/2011/07/read_my_lisp.html

Savin-Williams, R. C. (2006). Who's gay? Does it matter? *Current Directions in Psychological Science* **15**, 40–44.

Schramm, J. D. (2018). AI 'gaydar' could compromise LGBTQ people's privacy—and safety. *Washington Post*, 19 Feb. https://www.washingtonpost.com/opinions/ai-gaydar-could-compromise-lgbtq-peoples-privacy--and-safety/2018/02/19/172156bc-126d-11e8-9065-e55346f6de81_story.html

Shelp, S. G. (2003). Gaydar. *Journal of Homosexuality* **44**, 1–14.

Sides, J., and Citrin, J. (2007). European opinion about immigration: The role of identities, interests and information. *British Journal of Political Science* **37**, 477–504.

Srivastava, S. (2011). The precisely fuzzy science of gaydar. The Hardest Science blog, 21 Jul. http://hardsci.wordpress.com/2011/07/21/the-precisely-fuzzy-science-of-gaydar/

Tabak, J. A., and Zayas, V. (2012a). The roles of featural and configural face processing in snap judgments of sexual orientation. *PLoS One* **7** (5): e36671.

Tabak, J. A., and Zayas, V. (2012b). The science of 'gaydar.' *New York Times*, 3 Jun, p. SR10.

Toma, C. L., and Hancock, J. T. (2010). Looks and lies: The role of physical attractivenessd in online dating self-presentation and deception. *Communication Research* **37**, 335–351.

Valdes, F. (1995). Queers, sissies, dykes, and tomboys: Deconstructing the conflation of "sex," "gender," and "sexual orientation" in Euro-American law and society. *California Law Review* **83**, 12–33.

Valentova, J. V., Kleisner, K., Havlíček, J., and Neustupa, J. (2014). Shape differences between the faces of homosexual and heterosexual men. *Archives of Sexual Behavior* **43**, 353–361.

Wang, Y., and Kosinski, M. (2018). Deep neural networks are more accurate than humans at detecting sexual orientation from facial images. *Journal of Personality and Social Psychology*

114, 246–257.

Zytko, D., Grandhi, S. A., and Jones, Q. (2014). Impression management struggles in online dating. *Proceedings of the 18th International Conference on Supporting Group Work*, Sanibel Island, Fla., 53–62.