

Bayesian Advice for Gaydar-based Picking Up: Commentary on Lyons, Lynch, Brewer, and Bruno (2013)

Martin Plöderl

© Springer Science+Business Media New York 2013

Consider the following: About 5 % of people are homosexual. Is it possible to detect from appearance if somebody is homosexual? This seems to be possible. In a recent study, participants were able to detect if somebody was homosexual better than chance. In this study, participants had to judge the sexual orientation of individuals based on 80 photographs of faces: 40 photographs depicted a homosexual person; the other 40 photographs depicted a heterosexual person. It turned out that 70 % of the homosexual persons were correctly identified. Of the heterosexual pictures, only 20 % were incorrectly classified as homosexual.

Now the question: Assume that all individuals can be as accurate in detecting sexual orientation as in this study. If one sees a randomly chosen individual in the street and judges the person to be homosexual, what is the chance (between 0 and 100 %) that the person is indeed homosexual?

You may be surprised that the correct answer is a quite low probability: 16 %.¹ The problem described above is based on a recent article by Lyons, Lynch, Brewer, and Bruno (2013) who reported experimental research about women's accuracy of judging the sexual orientation of individuals based on photographs. The participating women could correctly identify gay/lesbian and heterosexual targets better than chance, i.e., the correct identification of gay/lesbian targets ranged between 58–65 %, and only 16–37 % of the heterosexuals were falsely identified as gay/lesbian. This adds to a number of other innovative experimental "gaydar" studies with comparable accuracy (e.g., Ambady, Hallahan, & Conner, 1999; Johnson, Gill, Reichman, & Tassinari, 2007; Rieger, Linsenmeyer, Gygas, Garcia, & Bailey, 2009).

But how does gaydar work outside of the laboratory? In the Lyons et al. study, accuracy of judgment was estimated by calculating the hit rate: the number of correctly identified homosexual targets (true positives) divided by the number of all presented homosexual targets. In addition, the false alarm rate was calculated by the number of heterosexual targets who were falsely categorized as gay/lesbian (false alarms or false positives) divided by all presented heterosexual targets.

When we want to translate the study findings into the real world, we have to consider the base-rate of gays/lesbians within the population, which is roughly at about 5 % and thus much less than in laboratory experiments where usually 50 % of the targets are gay/lesbian. By applying Bayes' Theorem, we can now calculate the probability of interest to us or to those who want to pick up potential same-sex partners from, let us say, a shopping-mall random-like sample; in other words, the probability that an individual is gay/lesbian, given our gaydar alarm rings.

$$\begin{aligned} \Pr(\text{gay}|\text{alarm}) &= \\ &= \frac{\Pr(\text{alarm}|\text{gay}) \Pr(\text{gay})}{\Pr(\text{alarm}|\text{gay}) \Pr(\text{gay}) + \Pr(\text{alarm}|\text{not gay}) \Pr(\text{not gay})} \\ &= 0.15 \end{aligned}$$

where $\Pr(\text{alarm}|\text{gay}) = .65$, $\Pr(\text{alarm}|\text{not gay}) = .20$, $\Pr(\text{gay}) = .05$, $\Pr(\text{not gay}) = 1 - \Pr(\text{gay}) = .95$

Based on the best findings by Lyons et al., chances are only at 15 % that gaydar alarms are correct. Therefore, gaydar errs way more often than not. This probability rises up to only

¹ It is not appropriate to use percentages for probabilities and this footnote could save me from being killed by my former statistics professor. However, people are used to percentages more than to probabilities ranging between 0 and 1. Moreover, in the context of the von Mises frequentist interpretations of probabilities, using percentages may be appropriate, i.e., the percentage of an event in an endless repetition of a random-generating process.

M. Plöderl (✉)
Sonderauftrag für Suizidprävention, Christian-Doppler-Klinik,
Ignaz-Harrer Str. 79, 5020 Salzburg, Austria
e-mail: m.ploederl@salk.at

18 % in the Rieger et al. (2010) study, where 81 % of targets were categorized correctly (assuming a hit-rate of 81 % and $100 - 81 = 19$ % false alarms). Even with a 90 % hit-rate and only 10 % false alarms, we still only have a probability of 32 % that a gaydar alarm is correct. Finally, with a 95 % hit-rate and 5 % false alarms, we can toss a coin if our gaydar is right. If the Lyons et al. results are applied to a population with 20 % gay/lesbians, the chance of guessing correct rises from 15 to 45 %. In probability samples, only around 1–2 % identify as gay/lesbian (Chandra, Mosher, Copen, & Sionean, 2011). Applying these base-rates, the probability of correct gaydar alarms drops down to 3–6 %. Figure 1 illustrates how strong the validity of gaydar depends on the base-rate of gays/lesbians in the population.

Of course, this does not downplay the importance of the highly innovative gaydar-research. However, it is very likely that the findings are misunderstood because readers may apply the findings directly to the real world. I presented the above problem in one of our staff-meetings consisting of 15 team-members who are predominantly psychiatrists, clinical psychologists, psychotherapists, and nurses. None of them was familiar with gaydar research. In the problem, I used the Lyons et al. results, but to increase ease of calculation, the assumed hit-rate was rounded to 70 % and the false alarm-rate to 20 %; the base-rate was given as 5 %. By Bayes' Theorem, the correct answer in the riddle is $\Pr(\text{gayalarm}) = 16$ %. The results indicate that there is a gross overestimation of this probability ($M = 48.36$, $SD = 28.81$), and only three participants gave probabilities smaller than .20. Moreover, one should keep in mind that the problem is easier to understand than reading the article because all necessary

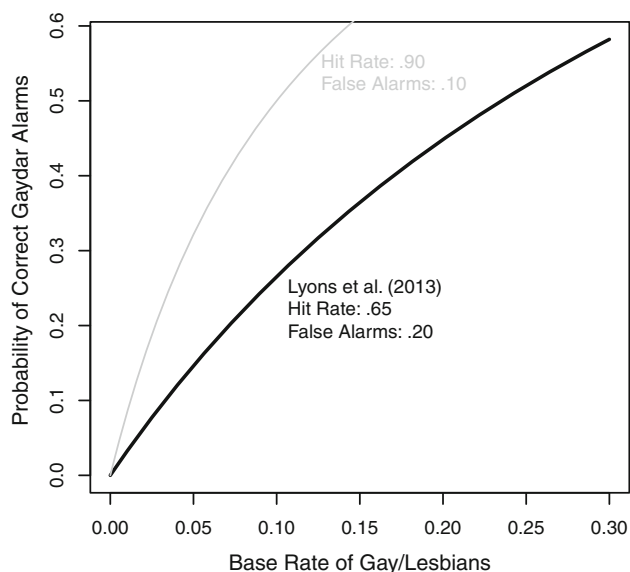


Fig. 1 Probability of correct gaydar alarms as a function of the base-rate of gay/lesbians in the population. The *thick solid line* is resulting from the Lyons et al. (2013) study (best results, homosexual raters, female targets). The *gray line* assumes an “ideal” study finding with a 90 % hit-rate and 10 % false-alarm rate

information is already extracted. The overestimation of gaydar-accuracy is not surprising because it resembles the well-known “base-rate fallacy” or “base-rate neglect.” From medical research, for example, it is known that even experts are frequently prone to the base-rate neglect, which can be highly problematic in the case of diagnosing rather rare events (HIV, breast cancer, prostate cancer, suicides, etc.) because, in such cases, it is common that there are substantially more false alarms than hits, even in the case of very sensitive and specific diagnostic tools (Hoffrage & Gigerenzer, 1998; Hoffrage, Lindsey, Hertwig, & Gigerenzer, 2000).

From what I am aware of, discussing the base rate-neglect is absent in gaydar research, and so this Letter was intended to counteract false or even unethical interpretations of the study findings. Furthermore, the fact that gaydar likely produces many more false-alarms than hits is also relevant from a public health perspective. Given that bullies use their gaydar to single out potential victims indicates that there are even more heterosexuals than actual gay/lesbians who experience homophobic violence associated with mental health problems. This was indeed found in one study (Reis & Saewyc, 1999, Table 4). Similarly, Tremblay, Plöderl, and Ramsay (2012) suggested expanding the homosexuality-factor in gay youth suicide research to also include youth who are assumed to be gay, independent of their actual sexual orientation, because, otherwise, the impact of homophobia is grossly underestimated.

Avoiding the base-rate fallacy is also crucial for interpretation of physiological assessments of sexual orientation (phallometry, vagometry, viewing time, eye-dilatation etc.). The estimation of the accuracy of such measures is often based on studies with a base-rate of homosexual individuals much higher than the population base-rate (e.g., Chivers, Seto, Lalumiere, Laan, & Grimbos, 2010). The same problems arise with physiological assessments of paraphilias (e.g., Seto, 2009), where, to my knowledge, a discussion of base-rate problems seems to be lacking.

Finally, researchers who do gaydar research or other sexual-orientation assessment research should always present hit-rates and false-alarm rates so that everyone can apply the laboratory findings to the real world with lower base-rates. Ideally, the researcher does this job for the reader, who may likely be unaware of his own base-rate fallacy.

To sum up, the Bayesian advice is that we likely err with our gaydar in natural settings. If one wants to do gaydar-based picking up, then one should choose an environment with many gay people (e.g., gaydar research laboratories).

References

- Ambady, N., Hallahan, M., & Conner, B. (1999). Accuracy of judgments of sexual orientation from thin slices of behavior. *Journal of Personality and Social Psychology*, *77*, 538–547.
- Chandra, A., Mosher, W. D., Copen, C., & Sionean, C. (2011, June 10). *Sexual behavior, sexual attraction, and sexual identity in the United*

- States: Data from the 2006–2008 National Survey of Family Growth. *National health statistics reports* (No. 36). Hyattsville, MD: National Center for Health Statistics. Retrieved from www.cdc.gov/nchs/data/nhsr/nhsr036.pdf.
- Chivers, M. L., Seto, M. C., Lalumière, M. L., Laan, E., & Grimbos, T. (2010). Agreement of self-reported and genital measures of sexual arousal in men and women: A meta-analysis. *Archives of Sexual Behavior*, *39*, 5–56.
- Hoffrage, U., & Gigerenzer, G. (1998). Using natural frequencies to improve diagnostic inferences. *Academic Medicine*, *73*, 538–540.
- Hoffrage, U., Lindsey, S., Hertwig, R., & Gigerenzer, G. (2000). Communicating statistical information. *Science*, *290*, 2261–2262.
- Johnson, K. L., Gill, S., Reichman, V., & Tassinari, L. G. (2007). Swagger, sway, and sexuality: Judging sexual orientation from body motion and morphology. *Journal of Personality and Social Psychology*, *93*, 321–334.
- Lyons, M., Lynch, A., Brewer, G., & Bruno, D. (2013). Detection of sexual orientation (“gaydar”) by homosexual and heterosexual women. *Archives of Sexual Behavior*. doi:10.1007/s10508-013-0144-7.
- Reis, B., & Saewyc, E. (1999). *Eighty-three-thousand youth: Selected findings of eight population based studies as they pertain to anti-gay harassment and the safety and well-being of sexual minority students*. The Safe Schools Coalition of Washington. Retrieved from <http://www.safeschoolscoalition.org/83000youth.pdf>.
- Rieger, G., Linsenmeier, J. A., Gygax, L., Garcia, S., & Bailey, J. M. (2010). Dissecting “gaydar”: Accuracy and the role of masculinity–femininity. *Archives of Sexual Behavior*, *39*, 124–140.
- Seto, M. C. (2009). Pedophilia. *Annual Review of Clinical Psychology*, *5*, 391–407.
- Tremblay, P., Plöderl, M., & Ramsay, R. (2012). Adolescent gay/lesbian suicide risk. An expanded homosexuality factor in adolescent suicide. Retrieved from <http://people.ucalgary.ca/~ramsay/adolescent-gay-lesbian-suicide-risk.htm>.