

Helping Behavior as a Subtle Measure of Discrimination Against Lesbians and Gay Men: German Data and a Comparison Across Countries¹

UTE GABRIEL²
University of Bern
Bern, Switzerland

RAINER BANSE
University of York
York, Great Britain

To unobtrusively assess attitudes toward lesbians and gay men, the wrong-number technique was used in a field experiment in Germany. The results are compared to studies using the same paradigm in Switzerland, Great Britain, and the United States. This approach gives a realistic picture of intercultural differences in social behavior against lesbians and gay men. Across studies, the results indicated that homosexuals are less likely to receive help than are heterosexuals. The variation of this effect between countries closely corresponded to the ranking of attitudes toward homosexuality assessed in survey studies. Contrary to survey studies, however, women showed only marginally less negative attitudes toward gay persons than men, when actual helping behavior was used as an attitude index.

There is a growing body of sociological and psychological literature dealing with the attitudes of heterosexuals toward gay persons or homosexual behavior (Baker & Fishbein, 1998; Fernald, 1995; Herek, 2000; Kite & Whitley, 1998; LaMar & Kite, 1998; Schope & Eliason, 2000). Survey as well as laboratory research has been conducted to study explicit opinions and evaluations, and their relation to other variables. One important determinant of attitudes toward lesbians and gay men has been identified in personality variables such as authoritarianism, religiosity, and sex stereotypes.

¹The authors are grateful to the German Telecom for their generous help and technical support for this research. We thank Meike Herting, Sebastian Korinth, Ina Parrhysius, Norman Radtke, as well as Susanne Schulze for running the German field experiment; Melanie C. Steffens and Ulrich Wagner for their helpful comments; and Iain Glen for critical comments and stylistic corrections.

²Correspondence concerning this article should be addressed to Ute Gabriel, Universität Bern, Institut für Psychologie, Unitobler, Muesmattstr. 45, CH-3012 Bern, Switzerland. E-mail: ute.gabriel@psy.unibe.ch

A further important factor is the national or cultural context as shown by the results of international surveys. For example, using data from the 1998–1999 International Social Science Survey Program, Kelley (2001) compared attitudes toward homosexuality in 29 countries. Based on a single-item measure (moral censure of homosexual behavior), the highest tolerance score was found for The Netherlands (77 out of a maximum score of 100), and the lowest for the Philippines and Chile (6 and 7, respectively).

In 1997, a sample of 9,818 citizens of the 15 countries in the European Union (aged 15 to 24) was interviewed for the Eurobarometer public-opinion surveys.³ Here, attitudes were assessed by asking about uneasy feelings when encountering gay men and lesbians in daily life. The proportion of respondents feeling uncomfortable with homosexual persons ranged from 23.8% in Northern Ireland to 6.1% in Spain (Melich, 2002).

Another variable that has been found consistently to be related to attitudes toward gay persons is gender. Whitley and Kite (1995) concluded from their meta-analysis of 66 studies that heterosexual men generally hold more negative attitudes toward gay persons than do heterosexual women. Furthermore, this effect seems to be moderated by the sex of the target: Men hold more negative attitudes toward gay men, whereas there seems to be no sex difference in attitudes toward lesbians (cf. Steffens & Wagner, 2004, who found women to hold more positive attitudes toward lesbians than men). With regard to sex differences, Kelley's (2001) study replicated the meta-analytic finding that men tend to be less tolerant, but the sex difference varied across countries. The highest difference was found in the Scandinavian countries (20-point difference), whereas in a few countries (including Russia, Chile, and the Philippines), male and female participants showed equally little tolerance for homosexual behavior. The study did not provide any information as to whether gay men and lesbians are viewed differently.

So far, the evidence based on self-report measures has indicated that attitudes toward homosexuals are influenced by culture, the sex of respondents, and the Culture \times Sex interaction. But what are the practical consequences of these findings? Can we conclude that lesbians and gay men are especially discriminated against by heterosexual men? And is it more likely that discriminating behavior will take place in Chile rather than in The Netherlands? Not necessarily, as there are at least two methodological problems that call into question the validity of survey data.

³Eurobarometer public-opinion surveys are conducted on behalf of the European Commission at least two times per year in all member states of the European Union. Since the early 1970s, they have provided regular monitoring of social and political attitudes in the European publics.

On the one hand, self-report measures can be confounded with self-presentation concerns as a result of social desirability, normative pressure, or the motivation to control prejudiced reactions (e.g., Dunton & Fazio, 1997; Gabriel, Banse, & Hug, 2003). Therefore, cultural or sex differences might reflect different sensitivities to political correctness rather than different evaluations. On the other hand, self-reported attitudes generally cannot be equated with overt behavior, as situational and personal features moderate the relationship (Kraus, 1995). Accordingly, self-reported positive attitudes do not necessarily imply that behavior will not be discriminatory; and similarly negative attitudes do not necessarily result in discriminating behavior, as demonstrated by LaPiere (1934) in his classic field study on attitudes and behavior toward the Chinese ethnic minority in the United States.

Given that the relation between attitudes and behavior is generally loose, and even more so in the context of prejudiced attitudes, it seems advisable to overcome the limitations of self-report measures by instead using unobtrusive behavior probes (e.g., a request for assistance) in order to get a more realistic picture of discrimination against specific minority groups. If behavior measures of prejudice reveal similar results as do survey methods, the latter could be used more confidently to estimate the amount of (positive or negative) discriminating behavior. Therefore, it is the aim of the present paper to examine whether the pattern of findings based on self-report measures of attitudes toward gay persons also can be found in research using overt social behavior toward lesbians and gay men.

To do this, we reviewed the literature, searching for field experiments using behavioral measures of prejudice toward homosexuals and heterosexuals. To exclude an influence of self-presentation concerns, we restricted our research to field experiments in which the participants were not aware of their behavior being observed. We conducted computer-based literature reviews in the psychological (PSYCLIT/PSYCINFO) and general (WEB OF SCIENCE) research periodicals through February 2003. Further, we reviewed references from these studies. In all, only eight studies were identified that fulfilled our requirements.⁴

Among these four studies (Ellis & Fox, 2001; Gabriel et al., 2001; Gore, Tobiasen, & Kayson, 1997; Shaw, Borough, & Fink, 1994), conducted in

⁴Seven further published studies used the lost-letter technique (Milgram, Milgram, & Harter, 1965), employing addressees of (fictitious) groups or organizations related to gay and lesbian topics (Bridges, 1996; Bridges, Anzalone, Ryan, & Anzalone, 2002; Bridges & Rodriguez, 2000; Bridges, Williamson, & Jarvis, 2001; Levinson, Pesina, & Rienzi, 1993; Waugh, Plake, & Rienzi, 2000). These studies were not included in our analysis because they studied community opinions about specific political issues (e.g., gay marriage, gay and lesbian teachers), rather than attitudes toward gay men and lesbians. Moreover, these studies did not feature a heterosexual control condition.

three different countries (Great Britain, Switzerland, United States), applied the wrong-number technique. In this experimental paradigm, households receive apparently wrong-number telephone calls that develop into requests for delivering a message to the actual addressee of the call (Gaertner & Bickman, 1971). In this paradigm, the formulation of the request implies the (minority-) group membership of the individual asking for help. For example, a request by a male person to pass on a phone call to his male partner implies that the petitioner is gay, while mentioning a female partner implies that the petitioner is straight. The amount of discriminating behavior against a minority then can be determined by comparing the number of positive responses with those obtained by majority members as a baseline measure of assistance (Gaertner & Dovidio, 1986). Our analysis will be based mainly on these four studies. The four other studies were excluded because of methodological shortcomings.

Helping behavior was also a dependent measure in two further studies conducted in Great Britain (Gray, Russell & Blockley, 1991; Tsang, 1994). Here, confederates wearing a T-shirt with either a pro-gay slogan or without any slogan approached shoppers on the street asking them to provide change for a 1-pound note. Although these studies also used a small request for help as the dependent measure, they had the methodological shortcoming that a blank T-shirt does not define a heterosexual control condition explicitly. Therefore, it remains open whether receiving more assistance in the blank T-shirt condition can be attributed to discrimination against lesbians and gay men or, alternatively, to discrimination against people wearing T-shirts with any political slogan.

In a study conducted in the United States by Hebl, Foster, Mannix, and Dovidio (2002), confederates wearing a cap with an imprinted slogan "Gay and Proud" or "Texan and Proud" applied for jobs at local stores. Indicators of formal (e.g., permission to complete a job application) and interpersonal (e.g., interaction length) discrimination served as dependent variables. Besides several additional problems, here again the control condition is problematic because effects may have been a response to lesbians and gay men, to Texan localism, or to any combination of the two.

In a study also conducted in the United States by Walters and Curran (1996) same-sex and opposite-sex couples (confederates) entered retail stores while an observer measured the time it took for them to be assisted by staff, and additionally recorded discriminatory behavior (e.g., being stared at, being pointed at). Contrary to the other studies, this study used proactive rather than reactive, spontaneous behavior as the main dependent variable. The couple did not approach the sales assistant; rather, it was up to the sales assistant whether and when they offered assistance. This behavior measure makes it difficult to compare this to the other studies.

As compared to the other paradigms, the wrong-number technique offers several advantages. The situation is relatively anonymous (as compared to face-to-face encounters), and it is easy to leave the situation (by hanging up) without any social consequences. Furthermore, there is an explicit request for aid (i.e., no situational ambiguity); and offering help is easy to do and does not take much time, effort, or other costs. Therefore, the monitored helping behavior is sufficiently frequent to provide a sensitive index of attitude toward the perceived social group and should be free of contamination by situational factors.

The straightforward manipulation of sexual orientation ensures that a homosexual condition is tested against a heterosexual condition, as opposed to nothing (as in Gray et al., 1991; Tsang, 1994) or against Texan localism (as in Hebl et al., 2002). As in the case of a questionnaire measure, the index behavior is prompted by the investigator. However, contrary to a survey situation, participants in helping experiments are not aware of being observed; and apart from the petitioner, there is no audience to be pleased or provoked. Thus, the behavior of interest—helping or not helping—should come very close to a pure indicator of discrimination against a specific minority group.

A further advantage of the wrong-number technique is that it allows for a straightforward and culture-fair comparison of studies conducted in different countries. In addition to investigating cultural differences in helping behavior toward homosexuals, this comparison makes it possible to test (a) whether the observed differences between countries correspond to those reported in the survey literature, and (b) whether the pervasive sex effect (i.e., that men hold more negative attitudes toward homosexuals than do women) can be replicated for the behavior measure.

We present a German field experiment using the wrong-number technique. Based on the results of previous studies, we expect homosexuals to receive less help compared to heterosexuals, and homosexuals to receive less help from male respondents than from female respondents. The obtained results will be compared to the results obtained in four previous studies conducted in Switzerland, Great Britain, and the United States, with reference to cross-national and sex effects.

German Wrong-Number Technique Study

Method

Sampling

Out of the 23 urban districts into which the city of Berlin is divided, we selected 5 districts with average social economic levels. The first three digits

of any telephone number are specific for the district (although in recent years, people may have kept their telephone numbers when they moved out of the district).

For use as callback numbers, German Telecom provided us with five telephone numbers (roughly corresponding to five city districts in Berlin) that were connected temporarily to a telephone located in a laboratory of the Psychology Department. Starting with these numbers, a list of eight-digit telephone numbers was generated by a computer program. The generated numbers varied either in the fourth and fifth or the sixth and seventh digit from one of the five callback numbers. The numbers were assigned in a pseudo-random fashion to the experimental conditions defined by the sex and alleged sexual orientation of the caller.

Phone calls were made on working days and on weekends between 6:00 p.m. and 9:00 p.m. during a 4-week period from January 2001 to February 2001. The caller rated the sex and age of the participants answering the phone. Participants who hung up before the request for help could be accomplished, who appeared to be younger than 18 years, who appeared not to be native speakers of German, or who presented themselves as representatives of public or private organizations were excluded. A total of 87 men and 118 women passed the exclusion criteria and were retained for analysis.

Design and Procedure

The design was a fully crossed 2 (Sex of Caller: male vs. female) \times 2 (Sexual Orientation of Caller: homosexual vs. heterosexual) \times 2 (Sex of Respondent: male vs. female) factorial design. The sex of the caller was established by a female or male caller who presented himself or herself either as "Anna" (female condition) or "Michael" (male condition).

Perceived sexual orientation was established by asking for one's romantic partner with a female name (*Lebensgefährtin Maria*) or a male name (*Lebensgefährte Peter*). The German word *Lebensgefährte* or *Lebensgefährtin* (translation is *mate, companion, or romantic partner*) unambiguously indicates both the sex of the partner and the type of relationship as a steady romantic relationship. Moreover, the sex of the caller and the sex of the romantic partner imply the sexual orientation of the caller.

According to a prepared script, the conversation of the male caller started by saying, "Hello, this is Michael speaking. May I talk to Peter [Maria]?" Typically, the respondent answered that there was no such person

and that the caller probably had dialed a wrong number. The caller continued

Oh, I'm very sorry, but listen. Now I'm in trouble because my car has broken down. I'm in a telephone box, and I wanted to tell my romantic partner [*Lebensgefährte/Lebensgefährtin*] that I will be late so he [she] won't worry. But now my phone card is nearly empty. I wonder if you could call him [her] for me and tell him [her] that he [she] doesn't have to worry? The number is ...

A positive reaction was coded if the respondent made the telephone call within the next 3 min (which was found to be a largely sufficient waiting time for incoming calls). A confederate who pretended to be a sibling of the target person took the call. The confederate said that he or she would pass on the message and expressed many thanks for the help of the caller. A response was scored as *no help* if the respondent refused to make the call, hung up after the caller had finished the prepared request for help, or if the call did not arrive in time.

Results

Overall, in 154 out of 205 calls (75.1%), help was provided. Table 1 shows the frequency with which help was provided to homosexual and heterosexual male and female callers, respectively. Heterosexual callers received significantly more help (83.5%) than did homosexual callers (67%), $\chi^2(1, N = 205) = 7.32, p = .01$. This difference was significant for male callers, $\chi^2(1, N = 105) = 4.03, p = .02$; and for female callers, $\chi^2(1, N = 100) = 3.33, p = .03$. Furthermore, male respondents, $\chi^2(1, N = 87) = 5.92, p = .01$, but not female respondents, $\chi^2(1, N = 118) = 1.25, p = .13$, provided significantly less help to homosexual callers.

Furthermore, the data pattern (cf. Table 1) suggests that male and female respondents discriminated against lesbian callers to the same extent, whereas gay men were discriminated against by male respondents only. To explore this three-way interaction effect, a logistic regression was conducted predicting helping behavior on the three main effects (sexual orientation, sex of caller, and sex of respondent), the three two-way interaction terms, and the three-way interaction term (computed as the cross-products of predictor variables). The contribution of each predictor was determined by using the likelihood ratio chi-square test (cf. Cohen, Cohen, West, & Aiken, 2003). The analysis revealed sexual orientation to be the only significant predictor, $\chi^2(1, N = 205) = 5.58, p = .02$. Thus, although there was a tendency for men to be less willing to pass on the call of gay persons than women, as shown in

Table 1

Frequency of Help by Sex of Caller and Sex of Respondent

| Perceived sexual orientation of: | Sex of respondent | | | | | |
|----------------------------------|-------------------|------|----------|------|----------|------|
| | Male | | Female | | Total | |
| | <i>n</i> | % | <i>n</i> | % | <i>n</i> | % |
| Male caller (<i>n</i> = 105) | | | | | | |
| Homosexual | 16 | 55.2 | 20 | 83.3 | 36 | 67.9 |
| Heterosexual | 15 | 83.3 | 29 | 85.3 | 44 | 84.6 |
| Total | 31 | 66.0 | 49 | 84.5 | 80 | 76.2 |
| Female caller (<i>n</i> = 100) | | | | | | |
| Homosexual | 14 | 63.6 | 19 | 67.9 | 33 | 66 |
| Heterosexual | 15 | 83.3 | 26 | 81.3 | 41 | 82 |
| Total | 29 | 72.5 | 45 | 75 | 74 | 74 |

the separate chi-square analyses, this difference did not reach statistical significance.

Comparison of the Five Wrong-Number Technique Studies

To compare the results of the five studies, the natural log of the odds ratios (ES_{LOR}), as well as an effect size based on the chi-square statistic (ES_r) were computed for each study. ES_r is defined as $\chi^2/N^{1/2} = \sqrt{N}$ and corresponds to the absolute value of the phi coefficient. The *odds ratio*, also referred to as the *cross-product ratio*, is an effect-size statistic that compares two independent samples in terms of the relative odds of an event. It is calculated as

$$(n_{11}n_{22})/(n_{12}n_{21})$$

defined in relation to a 2×2 table of observed multinomial frequencies (Fleiss, 1994). This kind of effect size is less common, but has statistical properties facilitating the combination of the effect sizes. The distributional form of the log-transformed odds ratio is approximately normal, with a mean of 0 and a standard deviation of 1.83. Thus, a negative value reflects a negative relationship, and a positive value reflects a positive relationship (cf.

Lipsey & Wilson, 2001). Unless otherwise noted, two-tailed p values are reported for the significance tests.

Comparison Across Countries

In the aforementioned survey study, Kelley (2001) obtained the following rank-order of decreasing tolerance points: Switzerland = 62; West Germany = 56; East Germany = 51; Great Britain = 46; and United States = 31. The Eurobarometer (Melich, 2002) surveying young Europeans provided no data on Switzerland or the United States. In the West German, British, and East German samples, 14.0%, 14.2%, and 18.8%, respectively, mentioned that they feel uneasy when confronted with gay persons.

Combining these results, Switzerland and the United States held the extreme positions of high and low tolerance, respectively; and Great Britain and Germany showed a similar intermediate level. It is important to note that this ranking is based on single-item measures, one of which (Kelley, 2001) focuses on homosexual behavior and not on homosexual persons. Such different types of judgments need not be consistent (cf. Kite & Whitley, 1998). Nevertheless, the measures used were identical in each country and, therefore, might serve as an estimate of cultural differences.

Table 2 presents the chi-square values and effect sizes of the five wrong-number technique studies, broken down by sex of caller. The effect sizes reflect how much the sexual orientation of the petitioner affected helping behavior. The studies are sorted in ascending order of the effect sizes for male callers (i.e., increasing discrimination against lesbians and gay men) from left to right. The resulting ranking of effect sizes replicates Kelley's (2001) findings for the male callers. For the female callers, only the positions of Great Britain and Germany are reversed.⁵

To quantify the cross-cultural heterogeneity of the effect sizes, we applied a meta-analytic test that is based on the Q statistic, which is chi square

⁵To further investigate this finding, the probability of a ranking corresponding to Kelley's (2001) findings by chance can be determined. There are five studies employing male callers, which can be sorted in 120 (5!) different rankings. Four out of the studies are consistent with the survey studies' ranking (i.e., rank order Switzerland, Germany or Britain, United States). The probability of obtaining a corresponding ranking by chance is 4 out of 120 (3.3%). For female callers, there are just 24 combinations (4!), with 2 of them (Switzerland; Germany or Britain; United States) corresponding to the results of the survey research. The probability of obtaining a corresponding ranking by chance is 2 out of 24 (8.3%), which is the lowest possible p value, given the number of studies. In sum, the correspondence of rank orders of countries obtained in survey research and using the wrong-number technique is significant for male callers, and marginally significant for female callers.

Table 2

Statistics and Effect Sizes of Homosexual Versus Heterosexual Callers by Sex of Caller

| | Male caller | | | | | Female caller | | | | |
|-----------------------------------|-------------|----------|----------|-----------------------|-------------------------|---------------|----------|----------|-----------------------|-------------------------|
| | <i>n</i> | χ^2 | <i>p</i> | <i>ES_r</i> | <i>ES_{LOR}</i> | <i>n</i> | χ^2 | <i>p</i> | <i>ES_r</i> | <i>ES_{LOR}</i> |
| Swiss study ^a | 80 | 0.07 | .79 | .03 | -0.14 | 80 | 0.67 | .41 | .09 | -0.45 |
| German study | 105 | 4.03 | .05 | .19 | -0.95 | 100 | 3.33 | .07 | .18 | -0.85 |
| British study ^b | 116 | 5.99 | .01 | .23 | -1.03 | 116 | 3.45 | .06 | .17 | -0.70 |
| U.S. study (Study 1) ^c | 80 | 18.34 | < .001 | .48 | -2.12 | | | | | |
| U.S. study (Study 2) ^d | 60 | 8.15 | .01 | .38 | -1.56 | 60 | 5.46 | .02 | .30 | -1.47 |

^acf. Gabriel et al. (2001). ^bcf. Ellis & Fox (2001). ^ccf. Shaw et al. (1994) ^dcf. Gore et al. (1997); there were no female callers in this study; LOR = natural log of odds ratios.

distributed with $k-1$ degrees of freedom, where k is the number of effect sizes. A significant result indicates that the variability of the observed effect sizes is larger than would be expected from sampling error alone. In this case, the observed effect sizes do not estimate a common population mean. In other words, there are differences among the effect sizes that have some source other than sampling error, such as differences associated with different study characteristics (Lipsey & Wilson, 2001). As the main difference between the studies analyzed here is the country where they were conducted, heterogeneity of the effect sizes most probably will reflect cultural differences in giving help to homosexuals.

The analysis shows that the effect sizes for female callers, $Q(3) = 1.52$, $p = .68$, are homogeneous. For the effect sizes for male callers, there is more heterogeneity across countries, but the effect reaches only marginal significance, $Q(4) = 7.93$, $p = .09$. These findings suggest that absolute differences in helping behavior toward gay and straight persons across countries are rather low. This makes it even more remarkable that we were able to replicate the ranking.

Gender Differences

According to the results of research using self-report measures of attitudes, men hold more negative attitudes toward homosexuals than do women (Whitley & Kite, 1995). This sex difference seems to vary across

countries. In Kelley's (2001) analysis, the sex difference is about 13 tolerance points in the United States, Switzerland, and Britain; 7 points in (ex-West) Germany; and 3 points in (ex-East) Germany.

With regard to the five wrong-number technique studies, we expected that the effect sizes for male respondents would be higher than those for female respondents. If the results of the behavior measure converge with those obtained in survey research, this difference should be less pronounced in the German study than in the remaining studies. In Table 3, the statistical results and effect sizes are presented by sex of respondent (U.S. Study 2 could not be included in this analysis because target gender was not reported).

The studies are ranked by the difference between male and female respondents' effect sizes (see Table 3, ΔES_{LOR} , with negative values reflecting more discrimination shown by males). The results partly confirm the expected pattern: The difference between male and female respondents was less pronounced in the German sample than in the American and the Swiss samples. In the Swiss sample, female respondents held a positive ES_{LOR} . In this sample, women actually helped homosexual callers more than heterosexual callers. An interesting but unexpected result emerged for the British sample: Here, male respondents hardly discriminated between homosexual and heterosexual callers; whereas, female respondents showed a (negative) effect size similar to the U.S. sample. But this result should not be overemphasized, as helpfulness in the British study was generally quite low (male callers = 29%, as compared to 55–77% in the remaining studies; female callers = 52%, as compared to 73–79%). Therefore, modest differences produced strong effects.

The distributions of the effect sizes were similarly heterogeneous for both sexes. The heterogeneity coefficient results were as follows: male respondents, $Q(3) = 8.01$, $p = .05$; and female respondents, $Q(3) = 7.69$, $p = .05$. Thus, male and female respondents reacted differently across the studies; therefore, mean ES represents the distributions less well.

Overall, male respondents, $M(ES_{\text{LOR}}) = -0.98$, $SE = .27$, $z = 3.69$, $p < .001$; and female respondents, $M(ES_{\text{LOR}}) = -0.71$, $SE = .24$, $z = 2.94$, $p < .001$, provided significantly less help to homosexual callers. The difference in mean effect sizes for male and female respondents is in line with expectations, but the effect was only marginally significant: $\Delta M = 0.27$, $SE(\text{pooled}) = .18$, $z = 1.48$, $p = .07$ (one-tailed). In the wrong-number paradigm, men's discriminating behavior against homosexuals was only marginally stronger than the discriminating behavior of women.

General Discussion

The aim of the present study was (a) to investigate the amount of antigay discrimination in Germany using an unobtrusive field experiment; (b) to

Table 3
Statistics and Effect Sizes by Sex of Respondent

| | Male respondent | | | Female respondent | | | Male vs. female respondent ^a | | | | |
|-----------------------------------|-----------------|----------|----------|-----------------------|-------------------------|----------|---|----------|-----------------------|-------------------------|-------------------|
| | <i>n</i> | χ^2 | <i>p</i> | <i>ES_r</i> | <i>ES_{LOR}</i> | <i>n</i> | χ^2 | <i>p</i> | <i>ES_r</i> | <i>ES_{LOR}</i> | ΔES_{LOR} |
| U.S. study (Study 1) ^b | 40 | 15.0 | .001 | .61 | -3.04 | 40 | 4.91 | .03 | .35 | -1.47 | -1.57 |
| Swiss study ^c | 80 | 3.67 | .06 | .21 | -1.10 | 80 | .62 | .43 | .09 | 0.42 | -1.52 |
| German study | 87 | 5.92 | .02 | .26 | -1.25 | 118 | 1.25 | .26 | .10 | -0.51 | -0.74 |
| British study ^d | 116 | 0.94 | .33 | .09 | -0.38 | 116 | 10.04 | < .01 | .29 | -1.22 | 0.84 |

Note. In the second U.S. study (Gore et al., 1997), sex of respondent was not assessed. ^a $\Delta ES_{LOR} = ES_{LOR}(\text{male respondent}) - ES_{LOR}(\text{female respondent})$. ^bcf. Shaw et al. (1994). ^ccf. Gabriel et al. (2001). ^dcf. Ellis & Fox (2001); the results are based on a reanalysis of the raw data, which were kindly provided by the first author of the original publication, J. Ellis. LOR = natural log of the odds ratios.

investigate sex differences in discriminating behavior; (c) to compare these results with data obtained in other countries; and (d) to evaluate whether behavior data obtained with the wrong-number technique are in accordance with attitude data collected in survey studies.

Discriminating Behavior in Germany

The empirical findings of the field experiment conducted in Berlin (Germany) impressively demonstrated that, in a real situation, gay persons could not count on the same helpfulness as could straight persons. Presenting oneself as gay significantly diminished the rate of help provided. As anticipated, men especially were less willing to pass on the calls of gay persons than those of straight persons.

This result is particularly striking, given that only a small favor was requested. Although an attitude of tolerance toward homosexuals seems to have developed in the last two decades in Western Europe (Rauchfleisch, 1995), the relation between straight and gay people seems to have not yet reached normality. On the basis of this study, inferences about the precise nature of the underlying psychological processes cannot be made. The discriminating behavior may be either a result of an active refusal to help a gay person, or a reluctance to get in contact with a gay person actively, even by phone. However, in daily life, this makes little difference for gay persons.

Discriminating Behavior Across Countries

Although the five field experiments conducted in the United States, Great Britain, Germany, and Switzerland did not differ significantly in the observed magnitude of discriminating behavior toward lesbians and gay men (as shown by the Q statistic), the effect size ranking of the experimental studies did replicate a ranking based on survey data (Kelley, 2001; Melich, 2002). This shows that there seem to be replicable cultural differences between the otherwise homogeneous discrimination rates.

With reference to the interaction of cultural differences and sex differences (i.e., men hold more negative attitudes than do women), the comparison across countries only partly replicated the ranking reported by survey research. In one of the four studies that were included, the direction of the sex difference was reversed: In the British study (Ellis & Fox, 2001), female respondents unexpectedly discriminated against gay persons more often than did male respondents. Regarding the sex of the respondent, the field experiments differed significantly in the magnitude of discriminating behavior. For female respondents, the effect sizes ranged from a slightly

positive discrimination of gay requesters in the Swiss study (Gabriel et al., 2001) to a medium effect of negative discrimination in the U.S. study (Shaw et al., 1994). For male respondents, the effect sizes ranged from a small effect of negative discrimination in the British study (Ellis & Fox, 2001) to a large effect in the U.S. study (Shaw et al., 1994).

As a possible limitation of the present meta-analysis, it must be noted that the five included studies were not strict replications of one another. First, the reason for not making another telephone call varied within the studies. In the British study (Ellis & Fox, 2001), the mobile battery was running out; in the German and Swiss study (Gabriel et al., 2001), the telephone card was nearly empty; and in the first U.S. study (Shaw et al., 1994), the caller had used his or her last quarter. In the second U.S. study (Gore et al., 1997) the reason was a further independent variable; namely, an urgency manipulation: The caller had no more change, or was down to the last quarter. As this manipulation had no significant influence on helping behavior (Gore et al., 1997), the two conditions were combined in our analysis.

Second, to identify the relationship as romantic, the two U.S. studies (Gore et al., 1997; Shaw et al., 1994) used the word *boyfriend* or *girlfriend*; Ellis and Fox (2001) used the word *partner*; Gabriel et al. (2001) used the word *Schatz* (translation is *darling*, *sweetheart*); and the present German study used the word *Lebensgefährte* or *Lebensgefährtin* (translation is *romantic partner*). In the studies of Shaw et al. and Ellis and Fox, the caller also mentioned that the couple was having a relationship anniversary that day. These slight methodological differences may have had an influence on the absolute level of helping in the specific study, but it is unlikely that they could account for gender or differences between countries in discrimination against lesbians and gay men.

In all five studies, people from urban or suburban regions were sampled in a similar procedure. Therefore, our results can hardly be explained by rural–urban differences in being helpful to less conformist people (e.g., Bridges, Ryan, & Scheibe, 1998). Furthermore, one could cite demographic and socioeconomic differences between the regions involved. The two U.S. studies, for example, took place in quite different regions (West Coast vs. East Coast; suburban environment vs. urban/rural transition environment), suggesting differences in the sociodemographic structure in the samples. Nevertheless, both studies yielded comparable results. At last, all studies used random samples of persons equipped with a telephone. It is very unlikely, therefore, that results were biased by systematic sampling error.

Comparing mean effect sizes for female respondents and male respondents across all countries, women tended to discriminate less against gay persons than did men, as predicted by research using self-report measures of attitudes (Whitley & Kite, 1995), but this sex difference did not reach

statistical significance ($p < .10$). This result is in line with field experiments not included in the comparison (Gray et al., 1991; Hebl et al., 2002; Tsang, 1994), which also did not show a significant sex effect.

When open, observable behavior is used as a measure of discrimination, women seem to differ less from men than in survey research. On the one hand, this finding might be a result of the fact that results of meta-analyses depend not only on the actual results of empirical research, but also on various decisions made while searching, selecting, and analyzing the relevant studies, as discussed by Oliver and Hyde (1995) with reference to gender differences in attitudes toward homosexuality (cf. also Oliver & Hyde 1993; Whitley & Kite, 1995). Therefore, gender differences simply might be overestimated.

But on the other hand, the diverging results indeed could reflect an interaction of gender by type of measure. It could be hypothesized, for example, that women do not report less negative attitudes because they are more tolerant, but because they are more motivated to present themselves in a favorable (e.g., politically correct) way. As self-report measures are more susceptible to presentational concerns than unobtrusive behavioral observations, self-report measures would reveal gender differences, but behavioral or other indirect measures would not (cf. Steffens, 2005, with reference to implicitly measured attitudes toward lesbians and gay men).

To summarize, a comparison of helping behavior in field studies and self-reported attitudes in survey studies reveals a similar rank order of countries. The finding of survey studies that women show less discriminating behavior than men, however, reaches only marginal significance in field studies. Furthermore, the interaction pattern of country and sex of respondent partly corresponded. The results from such different methodological approaches as nationwide surveys, attitudinal research using self-report measures, and field experiments designed to detect subtle discriminating behavior unobtrusively thus converged considerably; a finding that supports the often questioned validity of survey methods.

Because we did not compare attitudinal and behavioral measures of discrimination directly, the evidence presented here, of course, is inconclusive in this respect. Furthermore, we cannot judge the reliability either of the survey data or of the field experiments. However, we think that it is worthwhile to cross-validate survey data, especially by ecologically valid measures.

References

- Baker, J. G., & Fishbein, H. D. (1998). The development of prejudice towards gays and lesbians by adolescents. *Journal of Homosexuality*, 36, 89-100.

- Bridges, F. S. (1996). Altruism toward deviant persons in cities, suburbs, and small towns. *Psychological Reports, 79*, 313-314.
- Bridges, F. S., Anzalone, D. A., Ryan, S. W., & Anzalone, F. L. (2002). Extensions of the lost-letter technique to divisive issues of creationism, Darwinism, sex education, and gay and lesbian affiliations. *Psychological Reports, 90*, 391-400.
- Bridges, F. S., & Rodriguez, W. I. (2000). Gay-friendly affiliation, community size, and color of address in return of lost letters. *North American Journal of Psychology, 2*, 39-46.
- Bridges, F. S., Ryan, S., & Scheibe, J. J. (1998). Affiliation, community size, and destination in return of lost letters. *Psychological Reports, 83*, 1387-1393.
- Bridges, F. S., Williamson, C. B., & Jarvis, D. R. (2001). Differences in lost-letter responses from a seaside city. *Psychological Reports, 88*, 1194-1198.
- Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences*. Mahwah, NJ: Lawrence Erlbaum.
- Dunton, B. C., & Fazio, R. H. (1997). An individual-difference measure of motivation to control prejudiced reactions. *Personality and Social Psychology Bulletin, 23*, 316-326.
- Ellis, J., & Fox, P. (2001). The effect of self-identified sexual orientation on helping behavior in a British sample: Are lesbians and gay men treated differently? *Journal of Applied Social Psychology, 31*, 1238-1247.
- Fernald, J. L. (1995). *Interpersonal heterosexism* (pp. 80-117). New York: Guilford.
- Fleiss, J. L. (1994). Measures of effect size for categorical data. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 245-260). New York: Russell Sage.
- Gabriel, U., Banse, R., & Hug, F. (2003). *Predicting private and public helping behavior by implicit attitudes and the motivation to control prejudiced reactions*. Manuscript submitted for publication
- Gabriel, U., Beyeler, G., Däniker, N., Fey, W., Gutweniger, K., Lienhart, M., et al. (2001). Perceived sexual orientation and helping behaviour: The wrong number technique, a Swiss replication. *Journal of Cross-Cultural Psychology, 32*, 763-769.
- Gaertner, S., & Bickman, L. (1971). Effects of race on the elicitation of helping behavior: The wrong-number technique. *Journal of Personality and Social Psychology, 20*, 218-222.
- Gaertner, S. L., & Dovidio, J. F. (1986). The aversive form of racism. In J. F. Dovidio & S. L. Gaertner (Eds.), *Prejudice, discrimination, and racism* (pp. 61-89). New York: Academic Press.

- Gore, K., Tobiasen, M., & Kayson, W. (1997). Effects of sex of caller, implied sexual orientation of caller, and urgency on altruistic response using the wrong-number technique. *Psychological Reports, 80*, 927-930.
- Gray, C., Russell, P., & Blockley, S. (1991). The effects upon helping behaviour of wearing a pro-gay identification. *British Journal of Social Psychology, 30*, 171-178.
- Hebl, M. R., Foster, J., Mannix, L. M., & Dovidio, J. F. (2002). Formal and interpersonal discrimination: A field study of bias toward homosexual applicants. *Personality and Social Psychology Bulletin, 28*, 815-825.
- Herek, G. M. (2000). Sexual prejudice and gender: Do heterosexuals' attitudes toward lesbians and gay men differ? *Journal of Social Issues, 56*, 251-266.
- Kelley, J. (2001). Attitudes toward homosexuality in 29 nations. *Australian Social Monitor, 4*, 15-22.
- Kite, M. E., & Whitley, B. E. (1998). Do heterosexual women and men differ in their attitudes toward homosexuality? In G. M. Herek (Ed.), *Stigma and sexual orientation. Understanding prejudice against lesbians, gay men, and bisexuals* (pp. 39-61). Thousand Oaks, CA: Sage.
- Kraus, S. J. (1995). Attitudes and the prediction of behavior: A meta-analysis of the empirical literature. *Personality and Social Psychology Bulletin, 21*, 58-75.
- LaMar, L., & Kite, M. (1998). Sex differences in attitudes toward gay men and lesbians: A multidimensional perspective. *Journal of Sex Research, 35*, 189-196.
- LaPiere, L. T. (1934). Attitudes versus action. *Social Forces, 13*, 230-237.
- Levinson, K. S., Pesina, M. D., & Rienzi, B. M. (1993). Lost-letter technique: Attitudes toward gay men and lesbians. *Psychological Reports, 72*, 93-94.
- Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. Thousand Oaks, CA: Sage.
- Melich, A. (2002). *Eurobarometer 47.2OVR: Young Europeans; April-June 1997*. Cologne, Germany: Zentralarchiv für empirische Sozialforschung.
- Milgram, S., Mann, L., & Harter, S. (1965). The lost-letter technique: A tool of social research. *Public Opinion Quarterly, 29*, 437-438.
- Oliver, M. B., & Hyde, J. S. (1993). Gender differences in sexuality: A meta-analysis. *Psychological Bulletin, 114*, 29-51.
- Oliver, M. B., & Hyde, J. S. (1995). Gender differences in attitudes toward homosexuality: A reply to Whitley and Kite. *Psychological Bulletin, 117*, 155-158.

- Rauchfleisch, U. (1995). Homosexualität heute: Befreiendes coming out? [Homosexuality today: A releasing coming out?]. *Psychologie Heute*, 22, 66-69.
- Schope, R. D., & Eliason, M. J. (2000). Thinking versus acting: Assessing the relationship between heterosexual attitudes and behaviors toward homosexuals. *Journal of Gay and Lesbian Social Services*, 11(4), 69-92.
- Shaw, J., Borough, H., & Fink, M. (1994). Perceived sexual orientation and helping behavior by males and females: The wrong-number technique. *Journal of Psychology and Human Sexuality*, 6(3), 73-81.
- Steffens, M. C. (2005). Implicit and explicit attitudes towards lesbians and gay men. *Journal of Homosexuality*, 49, 39-66.
- Steffens, M. C., & Wagner, C. (2004). Attitudes towards lesbians, gay men, bisexual women, and bisexual men in Germany. *Journal of Sex Research*, 41, 137-149.
- Tsang, E. (1994). Investigating the effect of race and apparent lesbianism upon helping behavior. *Feminism and Psychology*, 4, 469-471.
- Walters, A. S., & Curran, M. C. (1996). "Excuse me, sir? May I help you and your boyfriend?": Salespersons' differential treatment of homosexual and straight customers. *Journal of Homosexuality*, 31, 135-152.
- Waugh, I. M., Plake, E. V., & Rienzi, B. M. (2000). Assessing attitudes toward gay marriage among selected Christian groups using the lost-letter technique. *Psychological Reports*, 86, 215-218.
- Whitley, B., & Kite, M. E. (1995). Sex differences in attitudes toward homosexuality: A comment on Oliver and Hyde (1993). *Psychological Bulletin*, 117, 146-154.