Check for updates

No magic bullet: estimating anti-immigrant sentiment and social desirability bias with the item-count technique

S. Rinken[1] · S. Pasadas-del-Amo[1] · M. Rueda[2] · B. Cobo[2]

© The Author(s) 2021

Abstract

Extant scholarship on attitudes toward immigration and immigrants relies mostly on direct survey items. Thus, little is known about the scope of social desirability bias, and even less about its covariates. In this paper, we use probability-based mixed-modes panel data collected in the Southern Spanish region of Andalusia to estimate anti-immigrant sentiment with both the item-count technique, also known as list experiment, and a direct question. Based on these measures, we gauge the size of social desirability bias, compute predictor models for both estimators of anti-immigrant sentiment, and pinpoint covariates of bias. For most respondent profiles, the item-count technique produces higher estimates of anti-immigrant sentiment than the direct question, suggesting that self-presentational concerns are far more ubiquitous than previously assumed. However, we also find evidence that among people keen to position themselves as all-out xenophiles, social desirability pressures persist in the list-experiment: the full scope of anti-immigrant sentiment remains elusive even in non-obtrusive measurement.

✉ S. Rinken
srinken@iesa.csic.es

S. Pasadas-del-Amo
spasadas@iesa.csic.es

M. Rueda
mrueda@ugr.es

B. Cobo
beacr@ugr.es

[1] Institute for Advanced Social Studies/Spanish Research Council (IESA-CSIC), Campo Santo de los Mártires 7, 14004 Córdoba, Spain

[2] Department of Statistics and Operations Research, University of Granada. Avda. Fuentenueva s/n, 18004 Granada, Spain

🙲 Springer

1 Introduction

Known technically as social desirability bias (SDB), the divergence of stated from true scores affects any survey on behaviors or attitudes that some interviewees are wary to disclose. Attitudes toward immigration and immigrants (ATII) are an intriguing case in point: the untold atrocities committed in the name of racial purity may induce survey respondents to choose evasive or dishonest answers whenever they perceive their views to be potentially interpreted as carrying ethnic overtones. And even regardless of such connotations, respondents might prefer to project an image of hospitality rather than voice grievances. Hence, unless SDB is eluded by specific survey techniques, unfavorable attitudes are prone to be underestimated by unknown margins.

Given this methodological challenge, one would expect ATII scholars to pioneer research on bias-reducing procedures. However, innovative survey methods such as the list-experiment, or item-count technique (ICT), have been employed sparsely in this field. Instead, much extant scholarship relies on expansive notions of prejudice as an alleged antidote against unrealistically low animosity estimates: any unfavorable opinion on international migration is commonly accepted as telltale of anti-foreigner prejudice. However, such dilution of the focal construct does not reliably detect hostile views. Equally troublesome, it imputes gratuitous hostility to people voicing potentially legitimate concerns. The combination of "false negatives" and "false positives" suggests that expansive notions of prejudice are methodologically flawed.

This study focuses on anti-immigrant sentiment (AIS), the affective core of xenophobia, as dependent and compares the estimates produced using a direct question with those obtained by an unobtrusive question format, namely, via ICT. We are thus able to:

(a) Quantify AIS and related SDB,
(b) Compare the predictor profiles of our two AIS estimators, and
(c) Pinpoint SDB covariates.

We use probability-based mixed-modes panel data comprising computer-assisted web interviews (CAWI) and computer-assisted telephone interviews (CATI). To control for potential mode effects, we include interview mode as predictor in logistic regression models for both (obtrusive and unobtrusive) animosity estimates.

The paper's structure is straightforward. We first derive our research hypotheses from a review of extant scholarship concerning SDB-reducing survey techniques in general and ATII research, in particular (Sect. 1). We then describe the dataset as well as the study's methodological constraints and choices (Sect. 2), present our findings (Sect. 3), discuss their implications (Sect. 4), and conclude (Sect. 5).

The data confirm some of our hypotheses while rejecting others. In line with expectations, the list-experiment originates significantly higher AIS estimates than the direct gauge, and especially wide margins of SDB are associated with respondent features such as better education, low social trust, higher age, and inactive labor market status, among others. Contrary to our expectations, predictors of both AIS gauges are found to coincide, CATI interviewing is not associated with lower AIS estimates as compared to CAWI, and the list-experiment is found to incur in discernible SDB among respondents keen to position themselves as all-out xenophiles.

These results suggest, firstly, that self-presentational concerns regarding the manifestation of AIS are strikingly pervasive. Secondly, our data show that while ICT is a promising

technique, it is not immune to social desirability pressures: even when employing ICT, the full scope of AIS remains elusive.

2 Literature review

2.1 Reducing SDB: a matter of privacy and anonymity

The increasingly ubiquitous digitization of all domains of life is opening up exciting new research options (Groves 2011; Hill et al. 2020). Recent developments such as affective computing and sentiment analysis (Cambria 2016) or automated hate-speech recognition (Greevy and Smiton 2004; Laaksonen et al. 2020) rely on digital traces, including social media usage, to detect sensitive attitudes such as AIS. While side-stepping traditional manifestations of response bias, such innovative data sources and research techniques are in turn vexed by various kinds of bias (Sen et al. 2019), and there are no agreed procedures (yet) for deriving population estimates from such data (Japec et al. 2015:872). As of today, for scholars aiming to estimate the prevalence of attitudes and behaviours in large populations, the self-report sample survey remains the foremost tool at hand (Groves 2011; Hill et al. 2020).

However, traditional survey methods are subject to manifold problems and limitations. Survey research relies on two main assumptions: the first one is that the sampled individuals are representative of the target population, and the second that respondents report information accurately (Groves et al. 2009). As every survey practitioner knows, accomplishing these goals is not a straightforward task as multiple sources of error may arise in the process. Some respondents may fail to understand the question or lack the information required to give a proper answer. And when questions are perceived as intrusive or embarrassing, respondents may deliberately distort their answers (Tourangeau and Yan 2007).

When faced with topics of a sensitive nature, some respondents will edit their responses in order to manage the impression they make on others or, arguably, even to deceive themselves (Paulhus 1984). The tendency "to make oneself look good in terms of prevailing cultural norms when answering to specific survey questions" (Krumpal 2013) is known as social desirability bias (SDB) or socially desirable responding and has been extensively studied by psychologists and survey methodologists. Extant research has shown that socially objectionable behaviors such as drug use, binge drinking, abortion and sexual risk-taking are usually underestimated in surveys, as are racism, sexism and other socially ill-regarded attitudes (cf. Krysan 1998; Tourangeau and Yan 2007; Krumpal 2013). SDB also explains why surveys tend to overestimate well-considered behaviors like voting, charitable giving, energy conservation, church attendance, seat belt use, and the like (Tourangeau and Yan 2007).

Mode comparison studies have consistently shown that self-administered modes of data collection usually yield more accurate answers to sensitive questions than interviewer-administered ones (Tourangeau et al. 2013; Tourangeau and Yan 2007) and that this effect is particularly strong for computerized forms of self-administration (Gnambs and Kaspar 2015; Richman et al. 1999). Yet, while privacy is generally accepted to be a necessary

condition when dealing with sensitive topics, some evidence suggests that it may not be sufficient to avoid SDB, depending on the perceived level of anonymity of the survey situation (Callegaro, Manfreda, and Vehovar 2015; Tourangeau 2018; Tourangeau and Yan 2007). Several studies reveal that SDB may still be an issue for self-administered surveys that ask for respondent identification (Joinson 1999), when survey notifications are personalized (Heerwegh et al. 2005; Joinson 1999),[1] when staff remain close by during Computer-Assisted Self-Interviews (CASI) (Liu and Wang 2016), or in panel contexts which rely on prior information on and communication with respondents (Coutts and Jann 2011).

To ensure that respondents perceive guaranteed anonymity, survey methodologists have developed specialized questioning techniques which "(make) it impossible to directly link incriminating data to an individual" (Nuno and Saint John 2015).[2] These techniques have been used successfully to inquire about different sensitive topics, from sexual risk behaviors or drug use to employee theft and vote buying (Aronow et al. 2015; Coutts and Jann 2011). Among them, the item count technique (ICT), also known as list experiment or unmatched count technique (Imai 2011), is gaining ground among scholars trying to quantify the effect of SDB on the measurement of sensitive behaviors and attitudes (Wolter and Laier 2014).

ICT randomly assigns respondents to two experimental groups, to then ask about the *number* of behaviors they have engaged in/abstained from or the *number* of attitudinal items they support/reject. The list proposed to the treatment group adds the sensitive item under research to the "innocuous" list offered to the control group. By computing the difference between the average numbers obtained for both groups, researchers can estimate the population proportion that supports (or rejects, as the case may be) the sensitive item net of social desirability pressures. The size of SDB is calculated by comparing these estimates to the proportion obtained with a direct question (DQ) regarding that same sensitive item (Lax, Phillips and Stollwerk 2016).

Numerous studies have found ICT to reduce SDB in self-administered paper questionnaires as well as in face-to-face, CATI, and CAWI surveys (Wolter and Laier 2014). However, some studies have found that even ICT may be subject to SDB and, as is the case with other survey instruments, to non-strategic respondent error (Ahlquist 2018). In some cases, ICT misreporting may be induced by a design that endangers the unobtrusive quality of the experiment: i.e. when none (floor effect) or all (ceiling effect) of the control items apply to a significant share of respondents (Blair and Imai 2012; Glynn 2013; Kuklinski, Cobb and Gilens 1997a, b). In other cases, respondents may remain suspicious of the instrument[3] and offer deflated (or inflated, in the case of desirable behaviors or attitudes) item-counts to send a clear signal of disassociation from (or association with) the sensitive item. Negative ICT estimates suggest the existence of a deflation effect, whereas estimates exceeding 1 indicate an inflation effect (Zigerell 2011); still, there is scant empirical evidence to date of such distortions.

Extant scholarship on SDB-reducing survey techniques suggests that the list-experiment generates higher AIS estimates than a direct question (H1). While there is no guarantee

---

[1] The evidence here is somewhat mixed as other studies have found that personalization does not induce social desirable responding (Gnambs and Kaspar 2015; Heerwegh 2005; Heerwegh and Loosveldt 2007).

[2] For a review of these techniques see Tourangeau and Yan (2007), Krumpal (2013) and Nuno and St. John (2015).

[3] In one study, approximately half of the participants declared not to trust the assertion that researchers were unable to derive individual answers from the experiment (Coutts and Jann 2011).

that unobtrusive question formats capture the true extent of AIS, the prevailing view in the extant literature is that ICT estimates lack discernible traces of SDB (H2).

2.2 Measuring AIS: is more always better?

Among ATII researchers, interest in innovative survey procedures such as ICT has been rather limited to date even though race-relations research, from which ATII students have borrowed numerous conceptual and methodological blueprints (Ceobanu and Escandell 2010; Fussell 2014), proved their worth (Gilens, Sniderman, and Kuklinski 1998; Heerwig and McCabe 2009; Kuklinski et al. 1997a, b; Kuklinski, Cobb, and Gilens 1997a, b; Redlawsk, Tolbert, and Franko 2010; Sniderman and Carmines 1997). Instead, most ATII scholarship confides in generic survey routines or diluted focal constructs as alleged antidotes against SDB. We first review these two kinds of studies, then those which do employ State-of-the-Art techniques. Table 1 provides an overview on extant measurement approaches and their application to ATII studies.

To keep response bias at bay, a sizable share of ATII research depends solely on generic survey routines such as confidentiality assurances, non-reactive interviewing, and non-suggestive semantics and scales. The possibility of dishonest answers is occasionally acknowledged and educational attainment flagged as a potential SDB covariate (e.g. Burns and Gimpel 2000:205), but for practical purposes, obtrusive ATII estimators are taken at face value. This run-of-the-mill approach prevails in studies aiming to explain migration policy preferences (e.g. Bohman and Hjerm 2016; Citrin et al. 1997; Hainmueller and Hiscox 2007; Hiers, Soehl, and Wimmer 2017; Sides and Citrin 2007), a thematic dimension on which large public-domain datasets, such as the European Social Survey, provide a nuanced range of head-on items. Even if unobtrusive indicators were available as readily, they are rather ill-suited for delivering dependents of explanatory models: the anonymity guarantee awarded by ICT and similar procedures comes at the price of severing any tie between individual respondents, on one hand, and scores of the sensitive item, on the other. This drawback was recently eased by the development of imputation techniques (Blair and Imai 2012; Chou, Imai and Rosenfeld 2017; Corstange 2009; Holbrook and Krosnick 2010; Imai 2011), but these entail high standard errors. Thus, from a model-optimization perspective, the aim of discerning ATII determinants is best served when all variables—including the dependent—originate in DQs. However, such models would be of limited value if respondents who candidly express unfavourable ATII were to differ substantially, in terms of sociodemographic and attitudinal profile, from those giving deceitful answers (Janus 2010). To assess this possibility, this study compares predictor models for obtrusive and unobtrusive gauges of the same attitude facet. We hypothesize predictors of ICT-based and DQ-based AIS estimates to differ at least partially from one another (H3).

Reliance on generic quality routines treats ATII as ordinary public preferences, i.e., favorable and unfavorable views are supposed to be equally legitimate. This assumption is dubious: the shockingly swift progression from "idle chatter" to two World Wars and the Holocaust forged a generalized commitment against *all* forms of ethnic and racial prejudice (Allport 1954:14–15)—including disrespectful verbalizations. To the extent to which unfavourable views are thought to convey ethnic or racial overtones, survey respondents may therefore shun their manifestation. Since such connotations are especially obvious with regard to outright animosity, researchers of anti-immigrant prejudice have recognized the need for safeguards against dishonest or evasive answers. Yet, what

Table 1 Overview: techniques for estimating sensitive attitudes and SDB

General category	Specific techniques (examples)	Benefits and advantages	Problems and limitations	Application to ATII research
Generic survey routines	Assurance of confidentiality Non-reactive interviewing Neutral questions and scales	Reduction of SDB Easy implementation Covariates analysis	Remaining SDB unknown	Very frequent, especially in studies of policy preferences
Indirect survey questions	Proxy indicators	Policy-related survey items widely available Easy implementation Covariates analysis	Dilution of focal construct False positives Magnitude of SDB unknown	Frequent in studies of AIS (cf. expansive notions of prejudice)
Self-administered surveys	Computer-assisted web interview (CAWI)	Reduction of SDB	Coverage bias in CAWI Remaining SDB unknown	Increasingly frequent
Survey experiments	Item count technique (ICT) Randomized Response Technique (RRT)	Potential for estimating both the sensitive attitude and, by comparison with a direct question, the extent of SDB	Difficult to implement Possibly still subject to SDB Limited use in statistical models	Few antecedents, most of which focussing on policy preferences
Digital traces ("big data")	Affective computing Pattern recognition	Real-time, low-cost Huge data volumes "Revealed preferences" (via online behaviour)	Requires data science skills Coverage and selection biases Access issues (proprietary data) Scant capacity of detecting covariates	Incipient

counts as a safeguard when true population parameters are unknown? Because undesirable attitudes cannot be validated externally, the highest estimator is generally accepted as best approximation (Höglinger and Jann 2018). Scholars of anti-immigrant prejudice have doubled down on the "more-is-better" approach by interpreting *any* unfavorable view regarding international migration as telltale of gratuitous hostility: "most theoretical models about attitudes toward immigration share the idea that anti-immigration attitudes are a form of prejudice" (Wilkes, Guppy, and Farris 2008:303). Acceptance of this conception was fueled by notions of "symbolic" or "subtle" prejudice (Gaertner and Dovidio 1986; Kinder and Sanders 1996; Sears 1988) and by the outright equation of perceived group-competition with prejudice (Bobo 1999; Quillian 1995); in contrast, classic formulations had considered such perceptions a (potentially forceful) trigger of prejudice, rather than its equivalent (Allport 1954:229–232). While inhospitable policy preferences may conceivably be less bias-prone than items regarding virulent animosity (Cea D'Ancona 2014), higher sample shares may also reflect nuanced positions toward distinct ATII facets (Ceobanu and Escandell 2010: 311–13). And while it is impossible to evaluate the justifications of natives' qualms (Esses, Jackson, and Armstrong 1998), the potential benefit of classifying any qualms as prejudice has to be weighed against the cost of conceptually eliminating the very possibility of legitimate concerns (Rinken 2016). Few studies on anti-immigrant prejudice (e.g. Hello, Scheepers, and Gijsberts 2002) employ specific gauges of animosity; instead, unwelcoming policy preferences or unfavorable impact assessments are used as indicators of "anti-foreigner sentiment" (Semyonov, Raijman, and Gorodzeisky 2006), "ethnic exclusionism" (Coenders and Scheepers 2003) or "xenophobia" (Hjerm 2007).

Such re-labelling overcomes none of the shortcomings of obtrusive measurement. As it happens, most experimental research on ATII[4] has focused on immigration control preferences, detecting sizable SDB and thus highlighting the inadequacy of expansive notions of prejudice as bias-reducing strategy. Janus' (2010) CATI-based study, conducted in 2005, reveals substantially more restrictionist preferences in ICT than direct measurement; this gap increases among liberal and well-educated respondents, suggesting that apparent pockets of tolerance derive from heightened propensity to bias. An (2015) also observes larger differences between direct and indirect measures of restrictionism among well-educated respondents. Comparing Janus' data with CAWI data for 2010, Creighton et al. (2015) detect more explicit opposition to immigration in 2010, whereas ICT results are similar; somewhat precipitously (since mode differences might play a role, cf. Dillman and Christian, 2005), they infer a time-trend of decreasing SDB. For their part, Creighton and Jamal (2015) find more overt opposition against naturalization of Muslim than Christian immigrants, whereas ICT yields similar results; they deduce added normative pressure to appear tolerant toward Christians. Similarly, Creighton et al. (2019) observe more *masked* opposition to racially similar immigrants than to racially different or poorer ones. As far as we are aware, just two papers address attitude facets other than policy preferences: Knoll (2013b) finds nativism in the US to be *over*-reported in direct measurement as compared to ICT, suggesting that associations with patriotism trigger inverse desirability pressures, while Krumpal (2012) employs the randomized response technique (RRT) to estimate

---

[4] A related yet distinct line of work examines implicit attitudes, i.e., well-established mental associations that tend to operate unintentionally (cf. Knoll 2013b; Malhotra, Margalit, and Mo 2013; Pérez 2010). In contrast, SDB occurs when respondents "feel pressure to misrepresent themselves … to appear more acceptable to others" (Knoll 2013a: p 1590).

xenophobia and anti-Semitism in Germany, obtaining modest increments vis-à-vis obtrusive measurement.

Internet-based data, such as social media and internet search data, are employed increasingly to study racist or anti-immigrant attitudes and their relation with populist communication strategies and right-wing voting (e.g., Stephens-Davidowitz, 2014; Heiss and Matthes, 2020). Arguably, such data elude traditional manifestations of SDB, even fomenting niche-specific social desirability dynamics that may favor explicit expressions of AIS. However, a combination of coverage and selection biases (Japec et al., 2015; Hill et al., 2020) makes such data unsuitable (as yet) for estimating AIS prevalence and covariates across large populations.

To resume, extant scholarship on ATII measurement suggests a fourth hypothesis: we expect AIS-related SDB to be associated with respondent characteristics that imply heightened susceptibility to normative pressures (H4). Specifically, we hypothesize significant gaps between obtrusive and unobtrusive AIS estimates, and hence SDB, among people with better education (H4.1), leftist ideology (H4.2), those interviewed in CATI mode (H4.3), and perhaps additional features (H4.4).

3 Methodology

3.1 Data and measurement

This study uses data from an ATII survey fielded in 2016[5] in the framework of PACIS, a probability-based mixed-modes panel run by the Spanish Research Council's Institute for Advanced Social Studies (IESA-CSIC).[6] PACIS comprises people aged 16 or more residing in private households in Andalusia, Spain's largest and most populous region. Stated immigration attitudes across Spain remained remarkably benevolent throughout the severe economic crisis initiated in 2008; the interrogative of why such an adverse context did not trigger increasing intergroup hostility is accentuated in Andalusia, where unemployment rates reached the eye-popping figure of 35% in 2013. Prior studies pinpoint a combination of dispositional and situational factors (Rinken 2016; Rinken & Trujillo 2018) but do not clarify to which extent anti-immigrant attitudes are masked in surveys. Spaniards' comparatively high proportions of nonresponse to immigration-related items of the European Social Survey (Piekut 2019) suggest that self-presentational concerns may play a relevant role.

The panel was recruited by off-line probability sampling and is conceived as a pool of respondents that are periodically invited to participate in different cross-sectional surveys. The survey on ATII targeted Spanish nationals only, achieving a 44.2% response rate[7] (n=1,232), 61% (n=753) of which via CAWI (default mode) and 39% (n=479) via CATI (backup mode). Non-response bias (Groves 2002; Groves et al. 2001) was corrected with raking ratio estimation weights based on official population statistics.[8] The questionnaire took about 18 min to complete on average (18.75 for CATI vs. 17.46 for CAWI).

---

[5] Fieldwork was conducted from June 7th through July 16th, 2016.

[6] See Domínguez Álvarez et al. (2020).

[7] AAPOR RR1.

[8] Three variables were considered for weighting: sex, educational attainment, and age group.

ICT Question (Q5)		Direct Question (Q13)
Some social groups cause sympathy while others do not. Please tell me toward how many of the following groups you rather feel antipathy. It does not matter which ones, just how many.	**Control group** Compulsive gamblers Overweight people Homeless people Bankers n= 422	How often have you felt sympathy for immigrants?
	Treatment group A Compulsive gamblers Overweight people Homeless people Bankers Immigrants n= 419	Very often Fairly often Sometimes Hardly ever → Direct AIS estimate Never n= 837

Fig. 1 Anti-immigrant sentiment ICT experiment design and question wording

The list experiment (Q5) was situated prior to any immigration-specific items, whereas the direct question was located in the questionnaire's final part (Q13). We used simple randomization without stratification to assign respondents to a control group and two treatment groups. The control group was asked toward how many among a list of four social groups (namely, compulsive gamblers, overweight people, homeless people, and bankers) they felt antipathy; a fifth group was added to treatment groups A (immigrants) and B (refugees). This work uses data from the control group (n = 422) and treatment group A (n = 419) (cf. Figure 1); AIS prevalence is estimated by comparing their mean "antipathy counts" (DiM estimator; cf. Imai 2011).

The list experiment's wording captures the emotional core of anti-immigrant prejudice (Allport 1954) quite literally. In contrast, since direct inquiries about outright antipathy are especially prone to evoke ethnic or racial overtones and, hence, elicit evasive or dishonest answers, our obtrusive AIS gauge asks for the antonym in order to somewhat relieve social desirability pressures: "How often have you felt sympathy for immigrants?" This question originally forms part of the "Subtle Prejudice Scale" (Pettigrew and Meertens 1995), which was devised on the assumption that prejudice is more often expressed by denying positive emotions than reporting negative ones. The two most unfavorable responses ("never" and "hardly ever") are classified as AIS, as opposed to three more positive options ("sometimes", "fairly often", "very often").[9] In prior research fielded in Andalusia, this item originated AIS estimates ranging from 16.1% (2011) to 11% (2013) (Rinken 2016); our dataset puts that number at 8.4% (2016).

3.2 Analytical procedures

Following recent recommendations (Ahlquist 2018; Blair, Chou, and Imai 2019), we rely on DiM estimators to quantify AIS and AIS-related SDB (cf. H1 and H2). To control for individual-level characteristics, Blair and Imai (Blair et al. 2019; Blair and Imai 2012; Imai 2011) have developed additional nonlinear least squares (NLS) and maximum likelihood (ML) estimators. We compute multivariate regression models (ML estimators) to infer the association between specific respondent characteristics and either AIS gauge (H3), on one hand, as well

---

[9] We excluded DK/NA (just 2.5%) from our analysis.

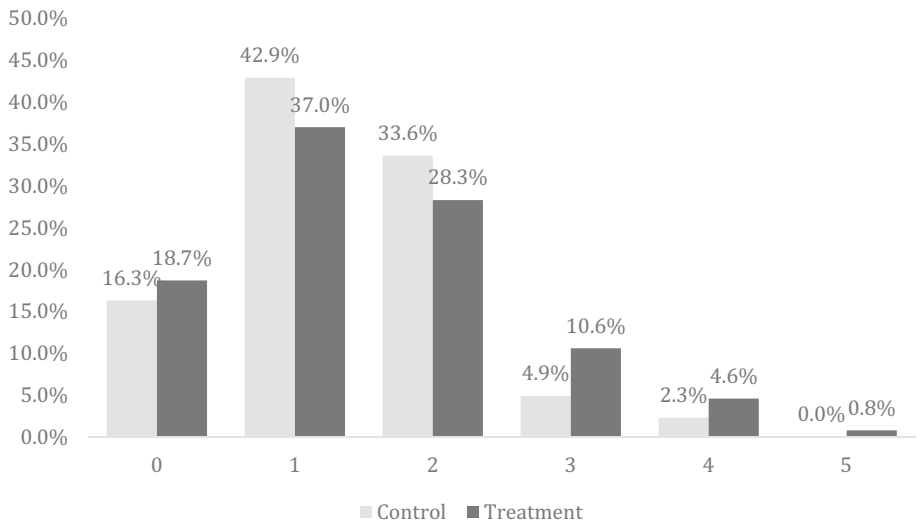| Table 2 Covariate balance (unweighted)	Treatment mean	Control mean	Difference
Sex			
Male	0.50	0.54	0.04
Female	0.50	0.46	− 0.04
Age			
18–34	0.25	0.23	− 0.02
35–49	0.32	0.31	− 0.01
50+	0.43	0.46	0.03
Education			
Up to primary	0.12	0.12	0.00
Secondary	0.55	0.55	0.00
Tertiary	0.33	0.33	0.00
Labor status	0.41	0.40	− 0.01
Unemployed	0.24	0.21	− 0.03
Economically inactive	0.35	0.39	0.04
Left–right self-identification			
Left (0–4)	0.38	0.39	0.01
Center (5)	0.39	0.37	− 0.02
Right (6–10)	0.23	0.24	0.01
Social class (self-identification)			
Upper/ Upper-Middle	0.08	0.08	− 0.00
Middle	0.54	0.53	− 0.01
Low/ Lower-middle	0.38	0.39	0.01

$+p < 0.1$; $*p < 0.05$; $**p < 0.001$; $***p = 0$

None of the differences shown in Table 2 are statistically significant ($p < 0.1$)

as the scope of SDB (H4), on the other. To model obtrusively measured AIS, we compute a standard logistic regression, whereas ICT-based AIS is modeled with the R-list package (Blair et al. 2016). Based on those models' regression coefficients, we obtain predictor-specific probabilities of declaring AIS in DQ and ICT measurement respectively; the difference between both AIS scores estimates SDB. This procedure pinpoints factors associated with SDB, net of other model variables (Blair and Imai 2012; Lax et al. 2016). In addition to sociodemographic items (sex, age, educational attainment, labor status, social class, and political ideology) and survey mode, we include two factors that, possibly due to SDB, mostly fail to yield significant impact in models regarding DQ-based AIS gauges: personal vulnerability to unemployment might plausibly spur AIS (Lancee and Pardos-Prado 2013), while higher levels of social trust should correlate inversely with AIS (Herreros and Criado 2009).[10] The process of estimation in the list experiment and the procedures and code used for analyzing its results are documented in Sects. 1 and 5, respectively, of the Online Appendix.

3.3 Evaluating the outcomes of our experimental design

Randomized experiments rely on the assumption that the treatment and control groups have similar covariate profiles. While some authors (Mutz 2011:108–12; Mutz and Pemantle 2015)

---

[10] See Online Appendix for question wording and coding.

Fig. 2 Observed data (weighted)

consider this condition to be met by the very randomization procedure, others (Gerber et al. 2014) insist on checking for any statistically significant differences. The unweighted covariate balance (Table 2) reveals no such differences, suggesting that randomization worked as intended in our experiment; to maintain that balance, weights were computed independently for each group ("within condition weighting"; Mutz 2011:119).

The validity of the ICT-ML model (Blair and Imai 2012; Imai 2011) relies on two additional assumptions. Firstly, control item counts must not differ depending on whether or not respondents are exposed to the sensitive item (Imai 2011). The R-list package includes a test to this avail (Blair et al. 2016); the Bonferroni-corrected p-value for our data (0.62) suggests no such design effect (cf. Online Appendix, Table A1). Secondly, ICT is supposed to always elicit truthful answers about the sensitive item; however, the "no liars" assumption (Imai 2011) may not be met when respondents perceive the anonymity of the experiment to be compromised. This scenario arises when a respondent's truthful score to all non-sensitive items is the same, thereby originating "ceiling" (all items) and "floor" (no item) effects respectively (Blair and Imai 2012; Glynn 2013). In our study, observed responses are distributed normally but somewhat skewed to the left, and the treatment group's proportion of zero items slightly exceeds the control group's (Fig. 2), suggesting potential floor effects. To evaluate their impact in our dataset, we have fitted different regression models accounting for the possible existence of ceiling and/or floor effects (Blair et al. 2018; Blair and Imai 2012). These cross-checks estimate the population proportions of dishonest respondents to range from 3.3% to 4.7% depending on the method employed (ML or Quasi-Bayesian approximation), and confirm that our results are robust to these distortions.[11]

---

[11] The Akaike Information Criterion (AIC) of our model (cf. Table 6) suggests that the weighted regression is to be preferred (AIC = 2034.9 versus 2080.2, cf. Online Appendix, Table A2). However, since weights are not implemented for the ceiling and floor models in the R-List package, these adjustments were computed with simpler models based on unweighted data (see Online Appendix, Table A3). The quality of

Table 3 Estimates of anti-immigrant sentiment (difference-in-means method)

	Control mean	Treatment mean	Difference in means	Direct question
Sample	1.301 (.042)	1.444 (.053)	0.143*	0.076 (0.009)
Weighted mean	1.340 (.043)	1.477 (.054)	0.137*	0.084 (0.01)
N	422	419		837

*$p < .05$

Table 4 ICT scores and DiM estimates by DQ response (regrouped categories)

Direct question	Control mean	Treatment mean	Expected Diff. in means	Actual Diff. in means	95% Conf. interval
Declare AIS	1.31 (0.14)	1.99 (0.24)	1.00	0.68 (0.28)*	(0.13, 1.23)
N	41	30			
Do not declare AIS	1.34 (0.05)	1.46 (0.06)	>0.00	0.12 (0.08)	(− 0.04,0.28)
N	379	387			

*$p < .05$

4 Results

H1: ICT originates higher AIS estimates than DQ Comparison between our two AIS gauges supports H1: while only 8.4% of respondents state AIS when asked directly, that proportion reaches 13.7% in the list-experiment (weighted data; Table 3). On account of the "more is better" approach, we deduce that ICT reduces SDB substantially: in relative terms, about 40% of prevalent AIS, as detected by ICT, goes unobserved in obtrusive measurement.

However, closer inspection reveals an interesting twist: 32% of respondents who declare AIS in DQ fail to do so in ICT (Table 4). Although the 95% confidence interval includes the predicted value of 1.00, this unexpected result warrants clarification. Rather than suspecting these respondents to have counter-factually declared AIS in DQ and answered correctly in ICT only ("inverse SDB", Lax et al. 2016:521ff.), we think that semantic nuances offer a more convincing explanation: our DQ-gauge classifies *lack of sympathy* as AIS, whereas the list-experiment refers to *antipathy*. Two possible interpretations arise: (a) some respondents who *lack sympathy* towards immigrants do not feel outright *antipathy*, or (b) the ICT's reference to *antipathy* triggers SDB among self-confessed (DQ) xenophobes. Since indifference seems plausible as an alternative rationale for lacking sympathy, we favor option (a). This interpretation suggests that our obtrusive gauge errs not only by missing some of the true xenophobes, on one hand, but also by incorrectly imputing AIS to some non-hostile respondents, on the other. "Subtle" DQ wording was meant to partially

---

Footnote 11 (continued)

the model based on unweighted data (AIC=2124.3) improves intangibly when we adjust for floor effects (AIC=2123.9), while model quality worsens slightly when adjusting for ceiling effects (AIC=2124.7).

Table 5 ICT scores and DiM estimates for specific DQ categories

Direct question categories	%	Control mean	Treatment mean	Difference in means	95% Conf. interval
Very often	33.3	1.32 (0.08)	1.10 (0.07)	− 0.22* (0.10)	(− 0.423;− 0.018)
n		137	141		
Fairly often	27.4	1.27 (0.08)	1.42 (0.10)	0.15 (0.13)	(− 0.107;0.409)
n		116	113		
Sometimes	30.8	1.44 (0.08)	1.82 (0.10)	0.38* (0.13)	(0.133; 0.637)
n		126	131		
Hardly ever or never[1]	8.5	1.31 (0.14)	1.99 (0.24)	0.68* (0.28)	(0.135; 1.225)
n	837	41	30		

*$p < .05$

[1]Response options merged due to low case numbers

elude SDB, yet ICT results highlight the drawbacks of that choice. For their part, 12% of respondents who declare some degree of sympathy toward immigrants in DQ are caught "red-handed" by ICT.

H2: There is no discernible SDB in ICT-based AIS measurement Even assuming that all interviewees who declare *lack of sympathy* in DQ score truthfully in ICT, evidence of SDB emerges from detailed ICT results for more benign DQ responses (Table 5). Unsurprisingly, DiM values decrease as DQ-sympathy increases, thereby illustrating the two gauges' correlation.[12] However, among interviewees who proclaim plentiful sympathy ("*very often*"), treatment group means are clearly inferior (-0.22) to control group means. Heeding Zigerell's (2011) deflation hypothesis, this finding suggests that some people mark artificially low ICT scores to preclude even the remotest possibility of being associated with AIS.[13] While the truthfulness of this subgroup's DQ scores remains uncertain, its ICT scores are demonstrably biased, thus distorting our overall AIS estimate: we reject H2.

H3: Predictors of ICT-based and DQ-based AIS estimates do not coincide Table 6 presents regression results (computed with the R-list package devised by Blair and Imai 2012) regarding the probability of scoring the list-experiment's sensitive item (first two columns), as well as the count of control items. Predictors coincide regardless of whether obtrusive (DQ) or unobtrusive (ICT) AIS gauges are employed: stated AIS increases discernibly among people with lower levels of social trust, centrist or right-wing political ideology, and inactive labor-market status, whereas sex, age, educational attainment, survey mode, social class, and perceived unemployment threat do not yield significant impact. H3 is rejected.

H4: SDB is associated with higher educational attainment (H4.1), leftist ideology (H4.2), CATI mode (H4.3) and perhaps additional features (H4.4) To detect SDB covariates,

---

[12] Among treatment group respondents, the count of antipathetic groups increases as sympathy decreases ($r_s = .300$, $p = .000$).

[13] Note that when considering various social groups (Q5), respondents ignore they will be asked explicitly about sentiment toward immigrants (Q13).

Table 6 Predictors of obtrusive and unobtrusive AIS measures (weighted)

AIC: 2034.87	ICT				Direct Question	
	Sensitive Item		Control items			
	Est.	S.E	Est.	S.E	Est.	S.E
(Intercept)	− 1.91	3.24	− 0.19	0.21	− 2.79	0.70***
Sex						
Male	1.11	1.24	0.07	0.09	0.36	0.30
Age						
18–34	− 2.36	1.61	− 0.13	0.12	0.54	0.39
35–49	− 2.02	1.33	− 0.04	0.12	0.44	0.40
Education						
Low/medium	− 1.41	1.04	0.07	0.09	0.18	0.30
Labor status						
Unemployed	0.47	1.66	0.27	0.12*	0.09	0.41
Inactive	2.49	1.50[+]	− 0.15	0.11	0.70	0.37[+]
Social class						
Upper/Upper-middle	2.14	1.37	− 0.17	0.20	0.71	0.46
Low/Lower-middle	− 1.09	1.23	0.14	0.09	− 0.03	0.33
Ideology						
Center	5.15	2.13**	− 0.12	0.10	0.72	0.35*
Right	3.84	2.00*	0.02	0.11	0.69	0.39[+]
Survey mode						
CAWI	1.62	1.32	− 0.30	0.10**	− 0.13	0.32
At risk of losing job						
Yes	2.27	1.43	− 0.04	0.11	0.11	0.35
Social trust (scale)	− 1.10	0.41**	− 0.07	0.02**	− 0.22	0.07**

[+] $p < 0.1$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$ (n = 782 for DQ model; n = 788 for ICT model)

again emulating Blair and Imai (2012), we use the coefficients obtained by both regression models to compute subgroup-specific differences between both AIS estimates. This works best for categorical variables, where significant differences highlight factors associated with SDB, *net of other covariates* (Table 7, last column).

We observe statistically significant magnitudes of SDB in a vast array of respondent categories, namely: people of either sex; older people (50 + years); people with secondary and tertiary education; those feeling threatened by unemployment and those who don't; the unemployed and the economically inactive; interviewees self-classifying as plain middle or upper-middle/upper class; and people with centrist or right-wing ideology. Albeit failing significance tests, sizable SDB is estimated for most other predictor categories: the only exception is leftist ideology. Since low cell counts impede meaningful significance tests for the model's continuous variable, we use plotted estimates as a stopgap (Fig. 3): higher levels of social trust are associated with lower levels of both AIS and SDB. Overall, our results highlight a striking pervasiveness of SDB across a broad range of respondent features.

Note that whereas our net-of-covariates AIS estimates reveal similar DQ scores for both survey modes, the ICT score is almost five percentage points higher in CAWI than CATI,

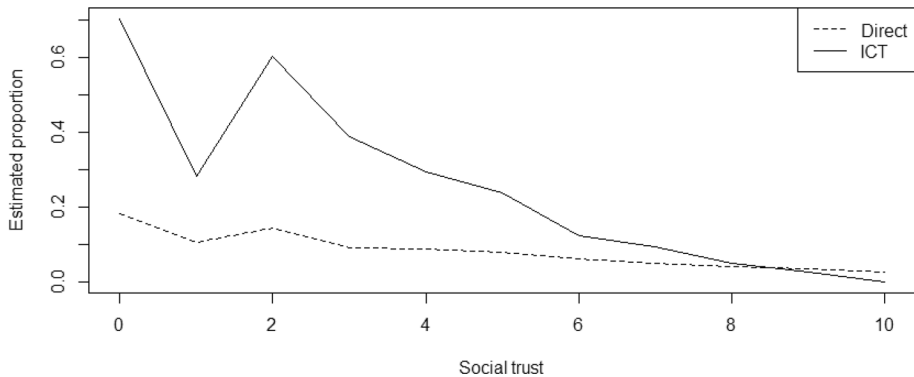**Table 7** Obtrusive and unobtrusive AIS estimates and SDB (weighted)

	DQ Estimate	ICT (ML)	SDB
Total	7.2	19.9	12.6**
Sex			
Male	8.0	21.7	13.6*
Female	6.3	17.8	11.5*
Age			
18–34	8.6	13.6	5.1
35–49	6.7	14.3	7.3
50+	6.6	28.0	21.4**
Education			
Up to secondary	7.7	15.8	8.1[+]
Tertiary	6.9	22.7	15.8*
At risk of losing job			
Yes	7.2	21.5	14.3**
No	7.3	18.1	10.8[+]
Labor status			
Employed	5.9	12.9	7.0
Unemployed	6.6	17.4	10.8[+]
Economically inactive	9.3	29.9	20.6**
Social class (self-identification)			
Upper/Upper-middle	13.7	44.9	31.2*
Middle	6.7	21.6	14.9*
Low/Lower-middle	6.6	12.5	5.9
Left–right (self-identification)			
Left (0–4)	4.8	4.3	− 0.5
Center (5)	8.8	34.2	25.4***
Right (6–10)	8.8	22.4	13.7[+]
Survey mode			
Web	6.9	21.6	14.7**
CATI	7.8	17.0	9.2[+]
Social trust[1]	[1]Estimates plotted in Fig. 3		

[+]$p < 0.1$; *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$

suggesting that the list-experiment works especially well in web-based self-administered questionnaires.

5 Discussion

Extant scholarship on attitudes toward immigration and immigrants (ATII) relies almost exclusively on obtrusive measurement; little is known about the scope and covariates of social desirability bias (SDB). Focusing on the emotional core of anti-immigrant sentiment

Fig. 3 Obtrusive and unobtrusive AIS estimates by level of social trust (weighted)

(AIS), this study employs two AIS estimates, one obtained by a direct question (DQ) and another one by non-obtrusive means (specifically, the item-count technique, ICT), in order to quantify AIS and related SDB, compare the predictor profiles of both AIS measures, and pinpoint SDB covariates. Four hypotheses were tested:

– H1: ICT produces higher AIS estimates than DQ.
– H2: There is no discernible SDB in ICT-based AIS measurement.
– H3: Predictors of ICT-based and DQ-based AIS estimates do not coincide.
– H4: AIS-related SDB is associated with better education (H4.1), leftist ideology (H4.2), CATI mode (H4.3) and perhaps additional features (H4.4).

Our findings suggest that the measurement of anti-immigrant sentiment is an even more treacherous endeavor than we had anticipated. As predicted (cf. H1), a sizable share of AIS goes undetected in obtrusive measurement: on aggregate, stated antipathy toward immigrants increased substantially (by 40% in relative and 5.3% in absolute terms) when employing ICT. However, our direct AIS gauge proved unreliable in an additional, unexpected way: one third of respondents who manifest *lack of sympathy* (DQ) do not rate immigrants as *antipathetic* (ICT). With hindsight, the pretended benefit of "subtle" DQ wording (cf. Pettigrew and Meertens 1995) was outweighed by the accrued drawback of mis-labelling indifference as antipathy, thereby inflating DQ-AIS and diminishing the ensuing SDB estimate. We therefore recommend future research to employ fully equivalent semantics and response options for obtrusive and unobtrusive estimators (Johnson 1998).

That said, this study's DQ format originated nuanced observations that would otherwise have been impossible: contrary to our expectations (cf. H2), ICT proved susceptible to SDB among people keen to position themselves (in the DQ) as all-out xenophiles. This subgroup's artificially low ICT score provides irrefutable evidence of deflation effects as described by Zigerell (2011). Somewhat ironically, by seeking to be "more Catholic than the Pope" these self-proclaimed xenophiles induce an underestimation of overall AIS prevalence. Further research is required to ascertain whether or not this limitation is particular to our study's panel-based design.

Comparison between ICT—and DQ-based regression models and ensuing estimates of AIS and SDB reveals that both AIS gauges are associated with the same predictors: animosity

increases tangibly among people with rightist or centrist ideology, low social trust, and inactive labor market status, whereas a range of additional factors fail to be significant in either model. Our expectation of varying predictor patterns for unobtrusive and obtrusive AIS gauges (H3) is thus refuted by the data. In contrast, our hypothesis regarding features associated with SDB is largely confirmed: SDB is associated not only with better educational attainment, but also with a wide array of other features. With regard to most predictor categories considered in our models, we discern statistically significant gaps between the magnitude of AIS as gauged by DQ and ICT, respectively. These gaps highlight a sufficiently broad variety of SDB covariates to suggest that self-presentational concerns regarding AIS are not limited to any particular part of the populace (cf. Krumpal 2013; An 2015); rather, they appear to be pretty much ubiquitous. We do not see how this important finding might originate in the "subtle" format adopted in this study for the obtrusive AIS gauge, given that a more aggressive DQ wording (in full analogy to the ICT's) would have yielded even lower estimates of AIS and, hence, wider margins of SDB; however, future research should cross-check this possibility. More research is also needed to ascertain whether such striking pervasiveness of bias applies to other countries as well, or else represents a distinctive feature of our observation arena.

There is one remarkable exception to the confirmation of H4: our data defy the expectation of people with leftist ideology being markedly SDB-prone (H4.2), instead detecting sizable bias among people with centrist or right-wing orientation. At face value, assuming truthful ICT-scores, these results contradict prior research on the relation between AIS-related SDB and political ideology (e.g. Janus 2010). Yet, given that ICT was in turn found to be affected by self-presentational concerns, it seems worth noting that people with left-of-center ideology account for a disproportionate share (42.5%) of respondents who declare all-out xenophilia in DQ; as reported, this DQ category's ICT scores are demonstrably deflated (Table 5). We are led to conclude that our non-obtrusive gauge underestimates the association of leftist ideology and SDB by an unknown margin.

Regarding mode effects, CAWI fails to reduce SDB in direct measurement notwithstanding the absence of social interaction. We assume this is due to concerns about anonymity in a panel-based survey (Callegaro et al. 2015; Tourangeau 2018; Tourangeau and Yan 2007); additional research is warranted on how to alleviate such concerns. In contrast, the combination of ICT and CAWI proves rather potent: *ceteris paribus*, controlling for a host of covariates, the list-experiment is found to work better when implemented by CAWI than CATI. Arguably, visualizing the list makes the experimental task easier (Lynn et al. 2012).[14] These findings complement extant scholarship on SDB in direct-interaction interviews (Krumpal 2013).

6 Conclusions

This study's results are ambivalent. On one hand, our data demonstrate that the list-experiment is a potent tool both for measuring sensitive attitudes, such as anti-immigrant sentiment, and for estimating the distortions caused by social desirability bias. This study reveals such bias to extend far beyond the sociodemographic groups flagged by the extant literature as susceptible in this regard. On the other hand, however, the data also show that

---

[14] Note that the count of non-sensitive items diminishes among CAWI respondents (Table 6, columns 3 and 4).

even this anonymity-maximizing survey technique is not immune against social desirability pressures. Ironically, the accuracy of our unobtrusive estimate of anti-immigrant sentiment is diminished by respondents aiming to prevent even the remotest doubt about their xenophile credentials. The data do not clarify whether such posturing is driven by concerns about anonymity protection in a panel-based setting, or else aimed at what is arguably the most important audience of all, namely, the respondents' own selves. However, our data demonstrate clearly that in future research, deflation effects have to be reckoned with.

Compliance with ethical standards

Conflicts of interest The authors declare that they have no conflict of interest.

Code availability (software application or custom code) Code is documented in the Online Appendix (Sect. 5).

References

Ahlquist, J.S.: List experiment design, non-strategic respondent error, and item count technique estimators. Polit. Anal. **26**, 34–53 (2018)

Allport, G.W.: The Nature of Prejudice. Addison-Wesley, Boston (1954)

An, B.P.: The role of social desirability bias and racial/ethnic composition on the relation between education and attitude toward immigration restrictionism. Soc. Sci. J. **52**, 459–467 (2015)

Aronow, P.M., Coppock, A., Crawford, F.W., Green, D.P.: Combining list experiment and direct question estimates of sensitive behavior prevalence. J. Surv. Stat. Methodol. **3**, 43–66 (2015)

Blair, G., Imai, K.: Statistical analysis of list experiments. Polit. Anal. **20**, 47–77 (2012)

Blair, G., Imai, K., Park, B., Coppock, A., and Chou, W.: list: Statistical Methods for the Item Count Technique and List Experiment. Available at the Comprehensive R Archive Network (CRAN), http://CRAN.R-project.org/package=list (2016)

Blair, G., Chou, W., Imai, K.: List Experiments with Measurement Error, Technical report. Princeton University, New Jersey (2018)

Bobo, L.D.: Prejudice as group position: microfoundations of a sociological approach to racism and race relations. J. Soc. Issues **55**, 445–472 (1999)

Bohman, A., Hjerm, M.: In the wake of radical right electoral success: a cross-country comparative study of anti-immigration attitudes over time. J. Ethn. Migr. Stud. **42**, 1729–1747 (2016)

Burns, P., Gimpel, J.G.: Economic insecurity, prejudicial stereotypes, and public opinion on immigration policy. Polit. Sci. Quart. **115**, 201–225 (2000)

Callegaro, M., Manfreda, K.L., Vehovar, V.: Web Survey Methodology. SAGE Publications, US (2015)

Cambria, E.: Affective computing and sentiment analysis. IEEE Intell. Syst. **31**(2), 102–107 (2016)

Cea D'Ancona, M.Á.C.: Measuring xenophobia: social desirability and survey mode effects. Migration Studies **2**, 255–280 (2014)

Ceobanu, A.M., Escandell, X.: Comparative analyses of public attitudes toward immigrants and immigration using multinational survey data: a review of theories and research. Ann. Rev. Sociol. **36**, 309–328 (2010)

Chou, W., Imai, K., Rosenfeld, B.: Sensitive survey questions with auxiliary information. Sociol. Methods Res. **49**(2), 418–454 (2017)

Citrin, J., Green, D.P., Muste, C., Wong, C.: Public opinion toward immigration reform: the role of economic motivations. J. Polit. **59**, 858–881 (1997)

Coenders, M., Scheepers, P.: The effect of education on nationalism and ethnic exclusionism: an international comparison. Polit. Psychol. **24**, 313–343 (2003)

Corstange, D.: Sensitive questions, truthful answers? Modeling the list experiment with LISTIT. Polit. Anal. **17**, 45–63 (2009)

Coutts, E., Jann, B.: Sensitive questions in online surveys: experimental results for the randomized response technique (RRT) and the unmatched count technique (UCT). Sociol. Methods Res. **40**, 169–193 (2011)

Creighton, M.J., Jamal, A.: Does Islam play a role in anti-immigrant sentiment? Exp. Approach, Soc. Sci. Res. **53**, 89–103 (2015)

Creighton, M.J., Jamal, A., Malancu, N.C.: Has opposition to immigration increased in the United States after the economic crisis? An Experimental Approach. Int. Migr. Rev. **49**(727), 756 (2015)

Creighton, M.J., Schmidt, P., Zavala-Rojas, D.: Race, wealth and the masking of opposition to immigration in the Netherlands. Int. Migr. **57**(1), 245–263 (2019)

Dillman, D.A., Christian, L.M.: Survey mode as a source of instability in responses across surveys. Field Methods **17**(1), 30–52 (2016)

Domínguez Álvarez, J.A, Pasadas del Amo, S., Sotomayor, R., Trujillo Carmona, M. and Rinken, S., Estudio sobre actitudes ante la inmigración y los inmigrantes en Andalucía (data file) 2020. DIGITAL. CSIC, https://doi.org/10.20350/digitalCSIC/12682 (2020)

Esses, V.M., Jackson, L.M., Armstrong, T.L.: Intergroup competition and attitudes toward immigrants and immigration: an instrumental model of group conflict. J. Soc. Issues **54**, 699–724 (1998)

Fussell, E.: Warmth of the welcome: attitudes toward immigrants and immigration policy in the United States. Ann. Rev. Sociol. **40**, 479–498 (2014)

Gaertner, S.L., Dovidio, J.F.: The Aversive Form of Racism. In Dovidio, J.F., Gaertner, S.L. (eds.) Prejudice, Discrimination and Racism, pp. 61–89. Academic, San Diego (1986)

Gerber, A., Arceneaux, K., Boudreau, C., Dowling, C., Hillygus, S., Palfrey, T., Biggers, D.R., Hendry, D.J.: Reporting guidelines for experimental research: a report from the experimental research section standards committee. J. Exp. Polit. Sci. **1**, 81–98 (2014)

Gilens, M., Sniderman, P.M., Kuklinski, J.H.: Affirmative action and the politics of realignment. Br. J. Polit. Sci. **28**, 159–183 (1998)

Glynn, A.N.: What can we learn with statistical truth serum? Design and Analysis of the List Experiment. Publ. Opin. Quart. **77**, 159–172 (2013)

Gnambs, T., Kaspar, K.: Disclosure of sensitive behaviors across self-administered survey modes: a meta-analysis. Behav. Res. Methods **47**, 1237–1259 (2015)

Greevy, E., Smeaton, A. F.: Classifying racist texts using a support vector machine. In: Proceedings of the 27th annual international ACM SIGIR conference on Research and development in information retrieval, pp. 468–469 (2004)

Groves, R.M.: Survey Nonresponse. Wiley, New York (2002)

Groves, R.M.: Three Eras of Survey Research. Public Opinion Quarterly **75**(5), 861–871 (2011)

Groves, R.M., Biemer, P.P., Lyberg, L.E.: Telephone Survey Methodology. Wiley, New York (2001)

Groves, R.M., Fowler, F.J., Couper, M.P., Lepkowski, J.M., Singer, E., Tourangeau, R.: Survey Methodology. Wiley, Hoboken (2009)

Hainmueller, J., Hiscox, M.J.: Educated preferences: explaining attitudes toward immigration in Europe. Int.l Org. **61**, 399–442 (2007)

Heerwegh, D.: Effects of personal salutations in e-mail invitations to participate in a web survey. Publ. Opin. Quart. **69**, 588–598 (2005)

Heerwegh, D., Loosveldt, G.: Personalizing e-mail contacts: its influence on web survey response rate and social desirability response bias. Int. J. Publ. Opin. Res. **19**, 258–268 (2007)

Heerwegh, D., Vanhove, T., Matthijs, K., Loosveldt, G.: The effect of personalization on response rates and data quality in web surveys. Int. J. Soc. Res. Methodol. **8**, 85–99 (2005)

Heerwig, J.A., McCabe, B.J.: Education and social desirability bias: the case of a black presidential candidate. Soc. Sci. Quart. **90**, 674–686 (2009)

Heiss, R., Matthes, J.: Stuck in a nativist spiral: content, selection, and effects of right-wing populists' communication on Facebook. Polit.l Commun. **37**, 303–328 (2020)

Hello, E., Scheepers, P., Gijsberts, M.: Education and Ethnic prejudice in Europe: explanations for cross-national variances in the educational effect on ethnic prejudice. Scand. J. Educ. Res. **46**, 5–24 (2002)

Herreros, F., Criado, H.: Social trust, social capital and perceptions of immigration. Polit. Stud. **57**, 337–355 (2009)

Hiers, W., Soehl, T., Wimmer, A.: National trauma and the fear of foreigners: how past geopolitical threat heightens anti-immigration sentiment today. Soc. Forces **96**, 361–388 (2017)

Hill, C.A., Biemer, P.P., Buskirk, T.D., Japec, L., Kirchner, A., Kolenikov, S., Lyberg, L.E. (Eds.): Big Data Meets Survey Science. A Collection of Innovative Methods. Wiley, Hoboken (2020)

Hjerm, M.: Do numbers really count? group threat theory revisited. J. Ethn. Migr. Stud. **33**, 1253–1275 (2007)

Höglinger, M., Jann, B.: More is not always better: an experimental individual-level validation of the randomized response technique and the crosswise model. PLoS ONE **13**, e0201770 (2018)

Holbrook, A.L., Krosnick, J.A.: Social desirability bias in voter turnout reports tests using the item count technique. Publ. Opin. Quart. **74**, 37–67 (2010)

Imai, K.: Multivariate regression analysis for the item count technique. J. Am. Stat. Assoc. **106**, 407–416 (2011)

Janus, A.L.: The influence of social desirability pressures on expressed immigration attitudes. Soc. Sci. Quart. **91**, 928–946 (2010)

Japec, L., Kreuter, F., Berg, M., Biemer, P., Decker, P., Lampe, C., Usher, A.: Big data in survey research: AAPOR task force report. Publ. Opin. Quart. **79**(4), 839–880 (2015)

Johnson, T.P.: Approaches to equivalence in cross-cultural and cross-national survey research. In: Harkness, J.A., Van der Vijver, F.J., Mohler, P.P. (eds.) Cross-Cultural Survey Equivalence. pp. 1–40. ZUMA, Mannheim (1998)

Joinson, A.: Social Desirability, Anonymity, and Internet-based Questionnaires. Behav. Res. Methods, Instrum. Comput. **31**, 433–438 (1999)

Joinson, A.N., Woodley, A., Reips, U.D.: Personalization, authentication and self-disclosure in self-administered Internet surveys. Computers in Human Behavior **23**(1), 275–285 (2007)

Kinder, D.R., Sanders, L.M.: Divided by Color: Racial Politics and Democratic Ideals. University of Chicago Press, Chicago (1996)

Knoll, B.R.: Assessing the effect of social desirability on nativism attitude responses. Soc. Sci. Res. **42**, 1587–1598 (2013a)

Knoll, B.R.: Implicit nativist attitudes, social desirability, and immigration policy preferences. Int. Migr. Rev. **47**, 132–165 (2013b)

Krumpal, I.: Estimating the prevalence of xenophobia and anti-semitism in Germany: a comparison of randomized response and direct questioning. Soc. Sci. Res. **41**, 1387–1403 (2012)

Krumpal, I.: Determinants of social desirability bias in sensitive surveys: a literature review. Qual. Quant. **47**, 2025–2047 (2013)

Krysan, M.: Privacy and the expression of white racial attitudes: a comparison across three contexts. Publ. Opin. Quart. **62**, 506–544 (1998)

Kuklinski, J.H., Sniderman, P.M., Knight, K., Piazza, T., Tetlock, P.E., Lawrence, G.R., Mellers, B.: Racial prejudice and attitudes toward affirmative action. Am. J. Polit. Sci. **41**(2), 402–419 (1997)

Kuklinski, J.H., Cobb, M.D., Gilens, M.: Racial attitudes and the 'New South.' J. Polit. **59**, 323–349 (1997a)

Laaksonen, S.-M., Haapoja, J., Kinnunen, T., Nelimarkka, M., Pöyhtäri, R.: The datafication of hate: expectations and challenges in automated hate speech monitoring. Front. Big Data (2020). https://doi.org/10.3389/fdata.2020.00003

Lancee, B., Pardos-Prado, S.: Group conflict theory in a longitudinal perspective: analyzing the dynamic side of ethnic competition. Int. Migr. Rev. **47**, 106–131 (2013)

Lax, J.R., Phillips, J.H., Stollwerk, A.F.: Are survey respondents lying about their support for same-sex marriage? Lessons from a list experiment. Publ Opin Quart **80**(2), 510–533 (2016)

Liu, M., Wang, Y.: Race-of-Interviewer Effect in the Computer-Assisted Self-Interview Module in a Face-to-Face Survey. Int. J. Publ. Opin. Res. **28**, 292–305 (2016)

Lynn, P., Hope, S., Jäckle, A., Campanelli, P., and Nicolaas, G.: Effects of Visual and Aural Communication of Categorical Response Options on Answers to Survey Questions, Technical report, Essex: ISER Working Paper Series (2012)

Malhotra, N., Margalit, Y., Mo, C.H.: Economic explanations for opposition to immigration: distinguishing between prevalence and conditional impact. Am. J. Polit. Sci. **57**, 391–410 (2013)

Mutz, D.C.: Population-Based Survey Experiments. Princeton University Press, Princeton (2011)

Mutz, D.C., Pemantle, R.: Standards for experimental research: encouraging a better understanding of experimental methods. J. Exp. Polit. Sci. **2**, 192–215 (2015)

Nuno, A., John, F.A.S.: How to ask sensitive questions in conservation: a review of specialized questioning techniques. Biol. Conserv. **189**, 5–15 (2015)

Paulhus, D.L.: Two-component models of socially desirable responding. J. Pers. Soc. Psychol. **46**, 598 (1984)

Pérez, E.O.: Explicit Evidence on the Import of Implicit Attitudes: The IAT and Immigration Policy Judgments. Political Behavior **32**(4), 517–545 (2010)

Pettigrew, T.F., Meertens, R.W.: Subtle and blatant prejudice in Western Europe. Eur. J. Soc. Psychol. **25**, 57–75 (1995)

Piekut, A.: Survey nonresponse in attitudes towards immigration in Europe. J. Ethn. Migr. Stud. (2019). https://doi.org/10.1080/1369183X.2019.1661773

Quillian, L.: Prejudice as a response to perceived group threat: population composition and anti-immigrant and racial prejudice in Europe. Am. Sociol. Rev. **60**, 586–611 (1995)

Redlawsk, D.P., Tolbert, C.J., Franko, W.: Voters, emotions, and race in 2008: Obama as the first black president. Polit. Res. Quart. **63**, 875–889 (2010)

Richman, W.L., Kiesler, S., Weisband, S., Drasgow, F.: A meta-analytic study of social desirability distortion in computer-administered questionnaires traditional questionnaires, and interviews. J. Appl. Psychol. **84**, 754 (1999)

Rinken, S.: Economic crisis and anti-immigrant sentiment: the case of Andalusia. Revista Española de Investigaciones Sociológicas **156**, 77–96 (2016)

Rinken, S., Trujillo, M.: The 'intergroup paradox' in Andalusia (Spain): an explanatory model. J. Ethn. Migr. Stud. **44**, 2392–2414 (2018)

Sears, D.O.: Symbolic racism, in eliminating racism: profiles. In: Katz, P.A., Taylor, D.A. (eds.) Controversy Perspectives in Social Psychology, pp. 53–84. Springer, Boston (1988)

Semyonov, M., Raijman, R., Gorodzeisky, A.: The rise of antiforeigner sentiment in European societies, 1988–2000. Am. Sociol. Rev. **71**, 426–449 (2006)

Sen, I., Floeck, F., Weller, K., Weiß, B., and Wagner, C.: A total error framework for digital traces of humans. https://arxiv.org/abs/1907.08228v2 (2019)

Sides, J., Citrin, J.: European opinion about immigration: the role of identities, interests and information. Br. J. Polit. Sci. **37**, 477–504 (2007)

Sniderman, P.M., Carmines, E.G.: Reaching Beyond Race, Cambridge, MA. Harvard University Press (1997)

Stephens-Davidowitz, S.: The cost of racial animus on a black candidate: evidence using Google search data. J. Publ. Econ. **118**, 26–40 (2014)

Tourangeau, R., Yan, T.: Sensitive questions in surveys. Psychol. Bull. **133**, 859 (2007)

Tourangeau, R., Conrad, F., Couper, M.: The Science of Web Surveys, Oxford. Oxford University Press, New York (2013)

Tourangeau, R.: Confidentiality, Privacy, and Anonymity, in The Palgrave Handbook of Survey Research, eds. D. L. Vannette and J. A. Krosnick, New York, NJ: Springer, pp. 501–507 (2018)

Wilkes, R., Guppy, N., Farris, L.: 'No thanks, we're full': individual characteristics, national context, and changing attitudes toward immigration. Int. Migr. Rev. **42**, 302–329 (2008)

Wolter, F., Laier, B.: The effectiveness of the item count technique in eliciting valid answers to sensitive questions. An evaluation in the context of self-reported delinquency. Surv. Res. Methods **8**(3), 153–168 (2014)

Zigerell, L.J.: You wouldn't like me when I'm angry: list experiment misreporting. Soc. Sci. Quart. **92**, 552–562 (2011)