Université de Lausanne
Les Cahiers de l'IEPHI
IEPHI Working Paper Series

Conformism in opinions about the welfare state in Switzerland

Results from a List Experiment

Lionel Marquis

N° 69 (2017)

Editeur responsable
Dr. Lionel Marquis

Université de Lausanne
Institut d'Etudes Politiques, Historiques et Internationales
Bâtiment Géopolis • 1015 Lausanne
CH – Switzerland

Tel +41 21 692 31 30
Fax +41 21 692 31 45
nicole.ferrari@unil.ch
http://www.unil.ch/iephi

GREC
(Groupe de Recherche sur les Élections et la Citoyenneté Politique)

La collection **Les Cahiers de l'IEPHI** vise à diffuser des travaux de chercheuses et chercheurs rattachés à l'Institut d'Etudes Politiques, Historiques et Internationales (IEPHI) de l'Université de Lausanne. Il peut s'agir de textes en prépublication, de communications scientifiques ou d'excellents mémoires d'étudiants. Ces travaux sont publiés sur la base d'une évaluation interne par deux membres de l'IEPHI. Les opinions émises n'engagent cependant que la responsabilité de l'auteur•e.

Les Cahiers de l'IEPHI sont accessibles gratuitement sur www.unil.ch/iephi (suivre le lien « Publications »).

The **IEPHI Working Papers Series** is intended to promote the diffusion of work in progress, articles to be published and research findings by researchers of the Institute of Political, Historical and International Studies, University of Lausanne. The papers submitted are refereed by two members of the Institute. The opinions expressed are those of the author(s) only.

The **IEPHI Working Papers** are available free of charge at www.unil.ch/iephi (click on « Publications »).

Layout : Nicole Ferrari

Couverture : Unicom, Université de Lausanne

Conformism in opinions about the welfare state in Switzerland
Results from a List Experiment

Lionel Marquis[i]
Institut d'Etudes Politiques, Historiques et Internationales, Université de Lausanne

Résumé

La désirabilité sociale est un biais de réponse, repérable dans toutes les méthodes d'interview, par lequel les individus cherchent à éviter l'opprobre frappant les opinions et comportements *socialement stigmatisés* en dissimulant leurs véritables préférences. Dans cette contribution, nous avançons que la désirabilité sociale concerne avant tout les personnes *ambivalentes* à propos d'un enjeu, ou celles qui ressentent *peu d'intérêt* pour celui-ci. Pour tester ces hypothèses, nous analysons les données d'une Expérience par Liste, dans laquelle des étudiants suisses ont été interrogés au sujet de leurs attitudes en matière de politique sociale. Une Expérience par Liste est une méthode permettant à certains sujets de s'exprimer librement, sans crainte d'être blamés pour avoir exprimé une opinion socialement "déviante", tandis que d'autres sujets restent conditionnés par l'influence normative du regard extérieur. Une comparaison entre ces deux groupes permet d'estimer la proportion d'individus « dissimulateurs » et d'identifier leur profil. Les résultats de l'expérience indiquent que 8 à 18 pour cent des sujets ont sans doute menti en répondant à trois questions : à propos de l'éligibilité aux prestations sociales, à propos de la discrimination à l'embauche, et de la manière la plus évidente à propos de l'équité fiscale. Les biais de désirabilité sont cependant plus prononcés au sein de certaines catégories : ils atteignent plus de 30 ou même 40 pour cent des étudiants avec un faible niveau de connaissance politique, mais également des étudiants plus jeunes et de ceux suivant un cursus en science politique — ce qui va largement dans le sens de nos hypothèses. Des analyses supplémentaires suggèrent que les individus ont une tolérance limitée pour la dissimulation ; lorsqu'on leur présente une liste de questions, ils semblent considérer qu'un "petit mensonge" est acceptable, mais pas davantage. Finalement, les promesses et les limites de notre approche sont discutées dans la conclusion.

Mots-clefs : désirabilité sociale, expérience par liste, attitudes, état social, Suisse

Abstract

Social desirability is a response bias affecting all interviewing methods, whereby people express opinions which do not correspond to their true preferences, in order to avoid being associated with socially *undesirable* opinions and behaviors. In this contribution, we argue that people are more likely to exhibit social desirability to the extent that they are *ambivalent* or have *little involvement* toward the issue at hand. To test these propositions we analyze the data from a List Experiment, where Swiss students were asked for their attitudes about welfare state policies. A List Experiment is an astute experimental design in which some subjects are induced to speak their mind freely, without fear of being blamed for having "socially deviant" opinions, while other subjects are left exposed to social normative influence. Simple between-group comparisons and recent refinements of the method allow to estimate the aggregate

[i] Maître d'enseignement et de recherche, Quartier UNIL-Mouline, Bâtiment Géopolis, CH–1015 Lausanne, lionel.marquis@unil.ch

proportion and the socio-political profile of "liars". The results of the experiment indicate that between 8 and 18 percent of subjects probably "lied" when answering about three items: about welfare entitlement (i.e., foreigners having the same rights to social services as Swiss citizens), about job nondiscrimination (i.e., of foreign against Swiss job seekers), and, most evidently, about tax fairness (i.e., more equity in tax policy, "even if I or my family were to be taxed more heavily"). Biases of this magnitude are relatively modest in comparison to other research (e.g., racial attitudes in the US). However, our findings indicate that social desirability responding (SDR) is more widespread among some categories represented in our sample of subjects. Most notably, SDR exceeded 30 or 40 percent of individuals with little political knowledge, of younger individuals, as well as of political science students, which is broadly consistent with our hypotheses. Additional analyses suggest that subjects tend to have a limited tolerance for lying; when confronted with a list of items, they seem to consider that "one little lie" is acceptable, but not more than that. Promises and limitations of our approach are discussed in the conclusion.

Keywords : social desirability, list experiment, attitudes, welfare state, Switzerland

Acknowledgements

This research was carried out at the University of Berne in the framework of a project on the "Political Consequences of Attitudes Toward the Welfare State" funded by the Swiss National Science Foundation, whose financial support is gratefully acknowledged (grant #100012–108274).

Table des Matières/Contents

1. Introduction[1]

This contribution explores whether and how social desirability affects public support for welfare state policies. Social desirability is the tendency of subjects to report attitudes or to exhibit behaviors which are at odds with their true preferences, in order to avoid being associated with socially *undesirable* opinions and behaviors. Put another way, social pressure and the fear of social exclusion lead to discrepancies between expressed ("overt") and sincere ("covert") opinions. As stated in more detail below, social desirability responding (SDR) has been established in many social and political domains. The question posed in this contribution is whether SDR also applies to welfare state policies, and whether SDR may be observed among university students, i.e., among individuals who are usually strong supporters of social protection.

In recent years, research has focused on the contextual and individual determinants of SDR. Despite the accumulation of studies, however, empirical evidence has remained scattered, largely inconsistent, and difficult to generalize across settings. Accordingly, the causes of individual differences in SDR are still poorly understood. In this paper, we argue that SDR can be broadly conceived as the consequence (or byproduct) of two common mental states or dispositions. First, we hypothesize that SDR is especially likely for people who have *conflicting views on an issue*. In the present case, citizens are often ambivalent toward the welfare state because their values and self-interest have conflicting implications for assessing the desirable scope and degree of state intervention. When forming their opinion on an issue, some people may, for example, let their material self-interest prevail over public interest and over a number of their own values. In conditions where overt answers are required, however, the same people are expected to give more weight to considerations which are consistent with the *perceived relevant social norm* on the issue. Second, we assume that, on average, SDR is more common among individuals with a *weak or diffuse political profile* — people with less integrated or less differentiated attitude systems, and less involvement toward the political issues of interest. The theoretical approach underlying these two general hypotheses will be described in sections 2 and 3.

To test our hypotheses, we use an unobtrusive method for detecting socially desirable responses (the "List Experiment"), which we present in section 4. Some two hundred students at a Swiss university were asked for their acceptance of specific arguments in favor of welfare state policies. Some of these arguments were deliberately chosen for their potential to elicit a conflict between material self-interest and "politically correct" values, that is, a *personal* conflict between anti-welfare and pro-welfare motivations. For example, we asked subjects whether they agreed with the statement that long-time foreign residents "should have the same rights to social services as Swiss citizens". To be sure, not every student will experience conflict when confronted with this question — arguably only a few will do, for example those sharing "chauvinistic" attitudes toward welfare benefits. But students form a community with a more or less agreed-upon set of norms and expectations. Therefore, it might be expected that *when* inner conflicts are aroused by such "sensitive questions" they will be resolved through compliance with community norms.

[1] This contribution is based on a re-analysis of experimental data collected in fall 2006 at the University of Bern, Switzerland, using a new multivariate method for analyzing List Experiment data that was not available at that time. The findings presented here thus differ (though not substantially) from those issued previously in technical reports and conference papers.

Our empirical results are presented in section 5. The experimental design and post-hoc analysis made it possible to assess the degree of "truthfulness" of opinions about *three statements* related to welfare state policies. To sum up, only a small proportion of subjects (typically less than 15 percent) are suspected of "lying" when answering questions about equivocal aspects of welfare policy. However, social desirability bias was found to be more pronounced among subjects with little political knowledge, among younger subjects, and among political science students — to name the most significant examples. This is broadly consistent with our "ambivalence" and "non-involvement" hypotheses.

From a methodological perspective, our results also make a contribution to SDR research in a number of ways. First of all, they show that the List Experiment method, originally conceived to explore the acceptance of "anger items" in large-N survey experiments, can be extended to ordinary opinion items (including focal items with very large acceptance rates) embedded in small-N experimental designs. Promises and limitations of our methodological approach are discussed in the conclusion of the paper.

2. Theoretical background

The literature on social desirability responding (hereafter **SDR**) points out that people sometimes hide their true feelings about an issue when they perceive that their feelings are at variance with a broad consensus on the issue. In many situations, "hiding one's feeling" is shorthand for "lying": people give fake answers to convey a positive self-image (e.g., Krysan, 1998; Berinsky, 2002; Krosnick, 1999). In this section, we provide a theoretical account of why these SDR tendencies exist and how they occur in the larger process of opinion formation.

2.1. Response editing

SDR may be conceived as a specific output of a larger process known as "response editing". When people are asked for their opinion on an issue, they do not simply retrieve an attitude about the issue from their long-term memory and verbalize it as accurately as they can. Instead, we assume that, most of the time, the so-called "attitudes" supposedly measured through standardized survey questions are in fact mere opinions formed "on the spot". Like other scholars, *we view attitudes not as fixed entities, but rather as transient, "temporary constructions"* (Zaller and Feldman, 1992; Wilson and Hodges, 1992; Wänke et al., 1996; Lord and Lepper, 1999; Schwarz and Bohner, 2001; Lord et al., 2004; Schwarz, 2007; Smith and Conrey, 2007; Yzerbyt and Kuppens, 2009). These constructions are derived from whatever considerations happen to be salient in people's mind at the moment they are asked for their opinion on some topic. To be sure, verbalized attitudes partly derive from the "true" underlying attitudes, values, or motivations to which they relate. But they are also heavily dependent on the many extraneous objects of consideration which are present in the broader response context. These objects are involved in an *editing process*, whereby attitudinal representations retrieved from memory are elaborated and adjusted to the incentives and requirements of the response context. This editing process includes "respondents' attempts to distort answers to avoid embarrassment, inconsistency, or other consequences" (Tourangeau et al., 2000: 256; see also Tourangeau and Rasinski, 1988).

Response editing occurs for many different reasons — for example, because respondents want to appear consistent in the answers they provide to related questions, because they strive for moderation and do not want to be considered too

partisan or extremist, or because they do not want to be *associated with socially undesirable attitudes and behaviors* (i.e., SDR). In other words, "social desirability operates as an editing process" (Holtgraves, 2004: 161) which tends to dissociate "edited" attitude self-reports from the more spontaneous responses made in the absence of social influence.[2]

Importantly, response editing — and SDR in particular — is not a "deviant" feature of exceptional individuals or situations. Rather, it is a pervasive phenomenon involved in most day-to-day processes of opinion formation. Accordingly, SDR research has strived to determine under what circumstances individuals adjust their responses to the perceived "opinion mainstream". It is generally admitted that SDR hinges on three types of factors: individual-level characteristics, item characteristics, and features of the response context.

2.2. Context and item factors

Depending on research goals, SDR can be conceived either as a fixed, built-in tendency, or as a more transient mechanism sensitive to the demands of a particular situation. This alternative has been conceived as a distinction between "response styles — biases that are consistent across time and questionnaires" and "response sets — short-lived response biases attributable to some temporary distraction or motivation" (Paulhus, 2002: 49). For example, the relevance of response sets varies with the degree of *anonymity of responses* and thus with the extent of "public embarrassment" created by socially undesirable answers (Paulhus and Reid, 1991). Accordingly, systematic differences in SDR have been found between different types of interviewing modes — including face-to-face interviews, self-administered computer surveys, Web surveys, and traditional CATI surveys (Holbrook et al., 2003; Presser and Stinson, 1998; Richman et al., 1999; Krysan and Couper, 2003; Brewer et al., 2004; Kreuter et al., 2008). Interestingly, it may be that publicity (as opposed to confidentiality) extends to situations in which responses are actually anonymous but are *perceived* to be socially engaging. Even totally anonymous, self-administered paper-and-pencil questionnaires may induce SDR. For what little "social interaction" there really is in such situations, it seems to be sufficient to elicit positive self-descriptions and the endorsement of socially approved opinions.[3]

---

[2] Holtgraves (2004) found that response latencies were increased under experimental conditions that elicited the greatest concern with social desirability. Likewise, social desirability is related to one important component of response latency in telephone surveys — sensitive survey questions such as about racial issues take a fraction of a second longer to answer than other questions (Johnson, 2004). This suggests that SDR is indeed a controlled "editing process" where individuals retrieve the relevant information from memory, evaluate it, and sometimes "adjust" it to social norms before responding (Tourangeau et al. 2000: chap. 9; Holtgraves, 2004; Tourangeau and Yan, 2007). Alternatively, SDR may proceed in an "automatic" or "stereotypical" way (Schaeffer, 1999: 118-120), consistent with the notion that many people are accustomed to "lying in everyday life" (DePaulo et al., 1996). Likewise, "explicit" and "implicit" attitude measures tend to become increasingly uncorrelated as the social sensitivity of issues grows and therefore social desirability concerns play a greater role in the expression of explicit attitudes (Nosek, 2005; Greenwald et al., 2009; Knoll, 2013).

[3] Under anonymous conditions, evidence of SDR has been found for self-reports of sexual behavior (e.g., Meston et al., 1998; Fisher, 2007), abortion (e.g, Jones and Forrest, 1992), personal hygiene (e.g., Gordon, 1987), racial prejudice (e.g., Sniderman and Carmines, 1997; Krysan, 1998; Krysan and Cooper, 2003; Berinsky, 1999, 2002), religious service attendance (Presser and Stinson, 1998), voting turnout (e.g., Clausen, 1968; Holbrook et al., 2003; Highton, 2005; Belli et al., 2006; Duff et al., 2007; Zeglovits and Kritzinger, 2014), voting for winners or incumbents (e.g., Wright, 1993; Mattei, 1998), and a whole array of social and political opinions and behaviors (e.g., Umesh and Peterson, 1991; Bradburn et al., 2004: chap. 3).

The fact that SDR occurs under anonymous conditions may sometimes be explained by the social norms that pervade a given community. As it was found, the degree of SDR varies with the larger *social and political context* of opinion formation and expression. Community pressure may arise even without a person's disclosing her views to anyone from that community, through feelings of guilt or betrayal (Bernstein et al., 2001). Thus, the greater the social pressure to display socially acceptable responses and behaviors in a given community, the larger the extent of SDR — overreporting of turnout in election surveys is a case in point (e.g., Karp and Brockington, 2005).

Beyond response context, item characteristics also play a role in SDR. For one thing, there is a gradation in the social sensitivity of issues — SDR is more widespread with respect to more sensitive issues. This is why both SDR and its predictors are largely issue-specific and situational, as meta-analyses of social desirability studies suggest (e.g., Richman et al., 1999; Tourangeau and Yan, 2007).

2.3. Individual-level factors

Individual determinants of SDR have first been studied in the field of personality psychology. In short, some individuals show a *chronic* disposition to SDR, while others do not (Paulhus, 1984, 2002; von Hippel and Trivers, 2011). Such dispositions need not concern us here, since it was not our objective to inquire into the personality foundations of SDR. In addition, the often-made distinction between conscious and non-conscious dimensions of self-enhancement is burred by the fact that the effect of social norms often goes unnoticed (Aarts et al., 2003; Nolan et al., 2008). Hence, individuals may provide socially desirable answers without conscious awareness of it.

More important for our present purposes, a number of socio-demographic or political attributes can be expected to play a role as *moderators* of SDR. (Moderators are variables for which SDR scores significantly vary between categories). Unfortunately, however, a great deal of empirical research on the moderators of SDR is inconclusive, suggesting that the effects of such moderators are specific to the situation and to the opinions or behaviors to be tested.[4] But there are other reasons why past studies have failed to yield meaningful generalizations about the moderators of SDR. To begin with, most of this research has overlooked the crucial fact that *individual-level variables are involved in several distinct mechanisms*. First, variables such as gender, age, ideology, or social class may influence people's *true preference* on an issue. This is obvious in the case of welfare state preferences because redistributive policies have structured society along socio-demographic and economic lines opposing welfare recipient and contributor groups. Second, these variables may differentiate individuals according to *how "sensitive" questions are perceived to be*. As we explain in fuller detail below, people with less involvement in political affairs (e.g., younger, non-politicized individuals) should find socially deviant opinions more "threatening" and thus more difficult to endorse publicly. Third, the same variables may be

[4] Purported effects of individual variables on SDR are often contradictory from one study to the next. For example, a meta-analysis suggests that age and gender are only modestly related to SDR (Ones and Viswesvaran, 1998; see also Belli et al., 2001; Mesmer-Magnus et al., 2006). However, other research finds no gender gap and suggests that age is confounded with education, which actually may show either a negative or a positive relationship with SDR (Klassen et al., 1975; Presser and Traugott, 1992; Bernstein et al., 2001; Heerwig and McCabe, 2009). In fact, all these relationships may vary according to the type of issue under consideration and to the type of SDR measure. For example, it has been found that higher education and middle-aged people are more prone to SDR with respect to the issue of electing a female US President, whereas women do not differ from men (Streb et al., 2008). Similarly, overreporting turnout in US elections was related either to highest or intermediate levels of political knowledge, political involvement, education, or age (Stocké and Stark, 2007; Belli et al., 2001).

important in determining *which social norm is relevant* for adopting "appropriate" opinions and behaviors in a given situation. For example, it may be of special interest to a pensioner to consider what other senior citizens think about some local issue; however, if the question at hand is how to reform the system of interregional fiscal transfers, the pensioner may turn to the population's opinion in her region. Accordingly, we envision a multiplicity of potential norms rather than a single social norm to which everyone would refer, and the role of individual-level variables in the three above mechanisms should not be considered in isolation from properties of the question posed and from characteristics of the context.

As the above analysis suggests, SDR may stem from complex interactions between characteristics of individuals, contexts, and items. We may now reformulate the idea that individual-level variables have several inputs into the SDR mechanism. On a purely logical basis, SDR is unlikely to occur if the respondent's true opinion equates with the opinion inferred from the most relevant available source of social influence. This is not to say that social influence does not matter in such cases, but rather that it may only *reinforce* existing opinions and hence is difficult to detect by conventional methods. Formally, this means that the probability that an individual-level variable affects SDR through "response editing" can be *independent* from the probability that it contributes to the formation of a "true" opinion in the first place — be it socially desirable or undesirable. Thus, for any group corresponding to the j^{th} value of a given variable, the *average group SDR estimate* should depend both on the group probability $p(D)_{ij}$ of disagreeing with the socially desirable item i, and on its probability $p(A \mid D)_{ij}$ of agreeing with the item, given initial disagreement, after some response editing mechanism has occurred: $\boldsymbol{p(SDR)_{ij} = p(D)_{ij} \times p(A \mid D)_{ij}}$. An empirical test of this two-tiered process is difficult with standard methods. Nevertheless it has heuristic value for hypothesis generation. In particular, it points to the possibility that moderator variables such as education or left-right ideology may have a *nonlinear* relationship with SDR, provided that they are inversely related to $p(D)_{ij}$ and to $p(A \mid D)_{ij}$.[5]

As noted above, one difficulty in assessing the role of moderators in SDR is that their effects are often highly dependent on the particular *context* (as defined by situation and/or item characteristics) under study. However, if we are to try to specify the role of each moderator in each context, we might not be able to cope with the amount of necessary explanation and ad hoc hypotheses. Rather than a piecemeal approach, what we need is an integrating framework enabling us to predict commonalities across the various "moderator × context" circumstances. For example, this framework should help us understand why certain moderator variables tend to have the same effect on SDR when operating in some types of situations and different effects in other situations. In the next section, we follow this path and formulate two hypotheses which are general enough to accommodate a great variety of circumstances *and* concrete enough to allow for non-trivial predictions.

[5] Ideology is a relevant example. Left-wing individuals are unlikely to endorse neoliberal and chauvinistic appeals to welfare state retrenchment. But those leftists who *do* support cuts in social spending, notably for self-interest reasons, may have especially ambivalent attitudes toward social security and may be quite susceptible to the normative influence of their ideological group. Conversely, right-wing individuals are expected to back retrenchment, at least to some extent. But few of these welfare state opponents will be ambivalent enough to be moved by uncongenial social norms. As $p(D)$ is almost certainly greater for right-wing subjects but $p(A \mid D)$ is probably greater for left-wing subjects, it is uncertain for which group the end estimate $p(SDR) = p(D) \times p(A \mid D)$ will be greater. Depending on how moderate left and moderate right citizens fare with respect to these probabilities, the overall relationship of ideological self-placement with SDR may have various functional forms, including U-shaped and inverted-U-shaped.

3. Hypotheses

Some people are, by disposition, more prone than others to engaging in conformist behavior. More often than not, however, SDR is a matter of circumstances. In this perspective, we argue that a potential for SDR arises when an individual's mental state tends toward ambivalence or non-involvement.

3.1. Ambivalence hypothesis (H1)

First, we argue that SDR is especially likely for people who have *conflicting views on an issue*, in other words, for people who are *ambivalent* on that issue. Indeed, a general lesson from research on racial attitudes in the United States is that Democratic Party supporters are conflicted between their liberal (egalitarian) attitudes and their existing racial prejudices, making them most vulnerable to SDR (e.g., Kuklinski et al., 1997; Krysan and Couper, 2006; Dovidio et al., 2009). In a similar vein, ambivalence may arise because a person's *values and self-interest* have different implications for the issue at stake.[6]

Importantly, attitudes about the welfare state — in particular about the implementation of redistributive policies — are deeply rooted in material self-interest *and* in key social values such as equality, fairness, and solidarity (e.g., Taylor-Gooby, 2001; Arts and Gelissen, 2001; Hinnfors, 2008; Calzada et al., 2014). Citizens endorsing egalitarian values tend to back proposals to ensure universal coverage and to reduce inequalities in income and life chances — whether or not they are net recipients of state support. Hence, due to the continuing pervasiveness of these values and lasting popular support for social policies, the welfare state has long proved surprisingly resilient to cutbacks in public spending (Pierson, 1994; van Kersbergen, 2000; Brooks and Manza, 2006; Castles, 2007).

However, in an era of permanent austerity and growing demographic pressure on state revenues, and more recently in times of severe economic crisis, these supportive values are increasingly challenged on two fronts. First, the rise of right-wing populist parties has promoted a new approach to the social role of the state known as "welfare chauvinism" (Banting, 2000; De Koster et al., 2013). In a nutshell, welfare chauvinists argue for giving priority to nationals in job recruitment and for restrictions in welfare eligibility that would exclude foreigners from a number of rights to social benefits. Second, the values conducive to welfare state support are challenged by the rise of *neoliberal or neo-conservative ideologies* underlying coordinated efforts to dismantle the welfare state (Korpi and Palme, 2003; Prasad, 2006). Importantly, neoliberal ideology puts a strong emphasis on the *maximization of one's interests*. As far as European welfare states were built on the basis of a "coalition of solidaristic interests" and "voluntary cooperation" between various political actors, social classes, and "risk groups" (Baldwin, 1990; Korpi, 2001), this principle of solidarity and risk pooling is likely to be undermined by the pursuit of narrow self-interest.

It is beyond dispute that the welfare state has created its own specific "clienteles" (e.g., pensioners, disabled people, or welfare-state employees; see Korpi, 2003). However, the stronghold of chauvinistic or neoliberal ideas in given constituencies may be expected to enhance the tendency of individuals to oppose whatever policies *fail* to provide (i) more personal benefits than personal costs; or (ii) more benefits to

[6] We focus here on this particular form of ambivalence, because we assume that values and self-interest are especially likely to have divergent implications for one's attitudes. To be sure, however, it is by no means the only type of ambivalence that can matter for welfare state attitudes.

nationals than to immigrants. Indeed empirical research suggests that self-interest is most successful in predicting "attitudes most directly and tangibly related to the welfare of the interested person" rather than "at higher levels of attitude abstraction" (Kumlin, 2007: 370-1; see also Kumlin, 2004). In particular, categories of voters may be driven by the prospect of beneficial tax cuts. In a related way, the strength of utilitarian motives is manifest in the greater popular support for programs aiming at improving security, relative to programs designed to achieve equality (Ross, 2000). Hence, policies such as guaranteed basic income will draw relatively little support if they can be framed as providing most benefits to "marginal" groups such as immigrants or disabled people (e.g., Bay and Pedersen, 2006).

Accordingly, we assume that *many individuals are fundamentally ambivalent toward social policy*. They are regularly exposed to competing arguments about the proper role of welfare institutions. In turn, these arguments enhance the memory accessibility of people's relevant dispositions, namely their value attachment to welfare policy goals *and* their material self-interest as welfare state contributors and recipients. Thus, citizens are likely to face any particular welfare reform proposal with *both* types of considerations in mind, and these may pull their opinions on the issue at hand *in opposite directions* (e.g., Mau, 2003: 23-27). For example, a house owner might support a new state-subsidized housing program to relieve families in need, but at the same time she might be reluctant to put additional burden on state finances to create benefits for which she is not eligible and which may entail tax increases. In fact, empirical studies of welfare state support have often considered the potential conflict between self-interest and values (e.g., Chong et al., 2001; Feldman and Steenbergen, 2001; Jaeger, 2006; Matthews and Erickson, 2008). This being said, *both* values and appeals to self-interest are reflected in the various social norms to which individuals may conform. To be sure, a strict separation of self-interest and values is not entirely justified, because it may be argued that *self-interest is itself a social norm*. Indeed it is such a powerful norm, at least in some contexts, that it contributes to its own confirmation and becomes self-fulfilling (e.g., Schwartz, 2012). According to one scholar, "the theory of self-interest has spawned a norm of self-interest, the consequence of which is that people often act and speak in accordance with their perceived self-interest solely because they believe to do otherwise is to violate a powerful descriptive and prescriptive expectation" (Miller, 2001: 194).

In cases of ambivalence, where values and self-interest give way to contradictory decision incentives, various individual and contextual circumstances may then determine which of the two types of dispositions will prevail. For example, self-interest may be especially likely to "crowd out" value-laden motives in certain types of welfare regime (Linos and West, 2003), in countries with low unemployment rates (Blekesaune and Quadagno, 2003), or within specific welfare recipient groups related to social class, age, gender, or income (Svallfors, 1997; Kulin and Svallfors, 2013).[7] Most importantly, ambivalent individuals can turn to social norms for assistance in evaluating the issue at stake. Specifically, social normative influence is arguably highest when individuals are ambivalent *and* suspicious that their opinions might be "overheard" or "leaked" to third parties. In such cases, individuals are likely to follow

[7] In addition to these structural factors, *situational* factors may determine the relative impact of self-interest and values on welfare state attitudes. For example, the specific form of questions targeting welfare support has been shown to "frame" the answers given by respondents. Support for redistribution is tremendous when questions are posed at a general, abstract level (e.g., "In society, the strong groups must care for the weaker ones"), but it plummets as soon as specific policies or recipient groups are mentioned (Kangas, 1997). Likewise, small changes in question wording can have a substantial impact on the measured level of welfare state support. Such "framing effects" are entirely consistent with our view of attitudes as *temporary constructions*.

the perceived relevant norm in order to avoid "public embarrassment" or feelings of guilt, and thus to escape fear of social exclusion.

As we alluded to above, an individual usually has several sources of normative influence, which so to speak compete for defining his or her social identity. Combining social desirability and social identity theories, we assume that individuals will pick different "reference groups" on different issues, depending on which of their multiple social identities are currently salient (e.g., Abrams and Hogg, 1990; Reicher and Hopkins, 1996; Wood, 2000; Levine et al., 2005).[8] Put differently, when *multiple* social identities and social norms are currently salient, and when these norms do not converge to the same definition of what is socially desirable, people may cope with cross-pressures by aligning with the group whose influence is perceived as most important or most diagnostic in determining the "socially appropriate" answer. Therefore, depending on specific individuals and situations, the activation of social norms will tend to trigger either specific values or personal material interests and thus to reinforce either pro-welfare or anti-welfare opinions. In other words, to become fully operational our ambivalence hypothesis has to be tailored to each specific situation.

In the present study, subjects were recruited among Swiss University students (see section 4). In this context, the dominant pro-welfare norm of the Swiss students' community may be so strong as to affect virtually every respondent — not only ambivalent respondents. To be sure, students who do not endorse egalitarian values *and* have vested interests against redistributive policies may be more likely express their opposition to such policies *when* the response context is perceived to be free from social constraints and sanctions. But many of them have probably developed social identities (e.g., through their membership in more intimate peer groups) which countervail the "dominant" norm in their community and prompt them to speak out their opinions irrespective of the response context. In contrast, individuals who are *ambivalent* about welfare policies are expected to conform to the perceived majority opinion — in this case, this would mean complying with the predominantly leftist and pro-welfare attitudes of the "campus community". Accordingly, *in this case*, ambivalent individuals should be more likely to support welfare policies in "public" settings than they would in more private conditions.

Overall, this "inner conflict resolution" hypothesis is consistent with the idea that SDR is a *response editing* mechanism. There is actually a body of empirical studies showing that both attitudinal ambivalence and minority status (i.e., the degree to which one's opinion is thought to deviate from the perceived social norm) lead to hesitation and discomfort in expressing one's opinions.[9] Drawing on our conceptual analysis of SDR, we propose to bridge the gap between these phenomena and assume that SDR is enhanced under conditions of attitudinal ambivalence.

---

[8] The term "reference groups" is used here in a generic sense to encompass all possible sources of social influence, ranging from primary groups and in-groups (e.g., based on profession, class, religion, or ethnicity) to mass media opinion and "mass opinion" at any conceivable level.

[9] For example, questions about issues that elicit *value conflicts* (e.g., women quotas in companies, an issue that tends to pit equality against merit) result in increased response times (Bassili and Fletcher, 1991; Sniderman and Carmines, 1997: 86-88). Likewise, issue opinions which are inconsistent with ideological orientations (as in the case of a liberal opposed to abortion rights) take more time to report than consistent opinions (Huckfeldt and Sprague, 2000). Interestingly, reaction times are also increased when respondents perceive their opinions to be at odds with those prevailing in their close interpersonal networks or in an important reference group (Huckfeldt and Sprague, 2000; Bassili, 2003; Huge and Glynn, 2013). Finally, an individual's willingness to express her political opinions has been shown to decrease with growing perceptions of a hostile "climate of opinion", but only when she is relatively uncertain about her opinions (Matthes et al., 2010). It could even be demonstrated that "faking" ambivalence is a frequently used *strategy* to avoid expressing one's opinion in a hostile opinion climate (Hayes, 2007).

3.2. Non-involvement hypothesis (H2)

Our second general hypothesis states that, on average, SDR is more common among individuals with a *weak or diffuse political profile*. By this we mean people with less integrated or less differentiated attitude systems, and less involvement toward the political issues of interest. Hence, social desirability is hypothesized to be more prevalent for people low in political knowledge, low in interest about the issues at hand, with less political experience, or with weaker or less salient political attitudes. One reason is that such people are less committed to (and less confident in) their own attitudes, less assertive in expressing their opinions, and less willing to defend a socially undesirable position.

For example, "defensive confidence" (i.e., an individual's self-perceived ability to counterargue when uncongenial information is encountered) has been shown to correlate positively with usual indicators of political involvement such as political participation, attention to politics, or political knowledge (Albarracin and Mitchell, 2004; Albarracin et al., 2012). More importantly, individuals who lack defensive confidence and political involvement tend to be uncomfortable with speaking their mind — they fear being negatively evaluated by others, and are reluctant to express themselves in public. This strongly suggests that socially unorthodox opinions are potentially more threatening to the self-image of people with less cognitive resources or less involvement in political affairs. To some extent, then, our second hypothesis is more detached from concrete political issues than the first, and it has more to do with "political personality". Thus it subscribes to the notion that social desirability is a manifestation of ego defense and represents an attempt at "impression management" (e.g., Paulhus and Reid, 1991).

Another reason why SDR is more prevalent for low-involvement individuals is that they lack strong and important beliefs about political issues, from which attitude statements can be automatically and/or effortlessly derived. Even those scholars who view attitudes as "stable and transparent" (van Harrenveld and van der Pligt, 2004; van der Pligt and de Vries, 1998) acknowledge that the role of the response context, which includes incentives to conform to social norms, is enhanced for the expression of attitudes based on weak beliefs or attributes. Conversely, individuals who hold intense and firm attitudes are more assertive of their opinions and more willing so express them in a straightforward manner (Neuwirth et al., 2007; Matthes et al., 2010). Likewise, people with more cognitive resources (such as well-educated and politically knowledgeable citizens) are better able to connect their abstract values to concrete social policies (Kulin and Svallfors, 2013). Therefore, their welfare state attitudes should be more dependent on values and less vulnerable to the effect of contextual cues.

3.3. Research question: Effects of item characteristics

Finally, to our hypotheses we add a *research question* concerning the effects of item characteristics. As we have argued above, the profile of people most susceptible to SDR will change according to the issues and to the way in which issues are framed in the questions. In other words, the "social sensitivity" of an issue is defined situationally, and the specific types of individuals to which our hypotheses apply will not be the same across issues. In assessing the role of issue questions, we will pay attention to the following features: the values to which issues appeal (e.g., equality vs. fairness); the main frame challenging these values (e.g., welfare chauvinistic or neoliberal); and the degree of concreteness and personalization of issues, as framed by the questions.

Using an experimental approach, our hypotheses will be tested with subjects recruited at the same time at the same university. This guarantees that the response context is, to a large extent, the same for all participants and that education level (known to affect both "private" and "public" opinions) is more or less constant. In turn this will increase our confidence in attributing variations in SDR to our focal individual-level variables, designed to measure ambivalence and political involvement, and to variations in item properties.

4. Measuring social desirability on welfare state issues

The goal of the experimental approach presented in this section is twofold. First, it is designed to observe whether and to what extent SDR manifests itself with respect to social spending and other welfare state issues. SDR on these issues may be less widespread than on more sensitive social problems such as race, homosexuality, or personal health (e.g., Berinsky, 2004). However, issue sensitivity is highly dependent on the context and on the framing of questions. As we suggest below, a substantial part of Swiss respondents should be caught in a dilemma between their commitment to pro-social values and the defense of their material interests, and these "cross-pressures" are expected to lead to SDR. Second, our experiment is tailored to identify those individuals most susceptible to SDR. To this aim, SDR estimates are computed for different subgroups of subjects (e.g., "low" vs. "high" political knowledge subjects). In total, ten variables will be measured to test particular aspects of the ambivalence and non-involvement hypotheses (see section 4.4).

4.1. Empirical ground: Swiss welfare policy

Like other states with a long democratic tradition, Switzerland was paradoxically late in introducing social security programs (Esping-Andersen, 1990: 14-16). As its development was both constrained and triggered by federalism, direct democracy, and religious factors (Armingeon et al., 2004; Manow, 2004; Obinger et al., 2005; Kriesi and Trechsel, 2008: chap. 10), the Swiss welfare state at the *national level* has lagged behind that of most other Western countries. Although Switzerland has been catching up with social-democratic and conservative regimes in the last decades, it was classified by Esping-Andersen (1990) in the "liberal" type of his famous threefold typology of welfare-state regimes (but see Trampusch and Mach, 2011). Swiss citizens' attitudes toward welfare are perhaps less deeply rooted in a long-standing commitment to social protection than is the case in other countries. These attitudes, in particular about policies that were recently introduced in the legislation (e.g., disability insurance, maternity leave), may be less resilient to attacks from neoliberal and populist quarters than they are in other countries.

But, more importantly, Switzerland is unique worldwide in submitting many proposals for welfare state reform to a *popular vote*. For example, no less than 29 referenda in the last hundred years (and half of them in the last 25 years) have been concerned with the development, regulation or financing of the Swiss pension system. These popular votes are golden opportunities for proponents of welfare state retrenchment to achieve immediate goals — stopping further expansion of welfare schemes or imposing cuts in social benefits. But these votes, as well as the campaigns preceding them, are also instrumental in the long run, as they put the idea of retrenchment on the political agenda and challenge established social rights by questioning their "irreversible" character.

Retrenchment as a *legitimate political question* will necessarily imbue popular attitudes toward the welfare state. This is reflected most directly in the substantial number of retrenchment and expansion proposals which have been respectively accepted and refused by majorities of Swiss citizens in popular votes. However, consistent with the "compensation strategy" outlined by Pierson (1994), the most successful retrenchment proposals have been the ones that combined retrenchment and expansion elements, while proposals "lacking such an expansionary component have been systematically rejected by voters" (Obinger et al., 2005: 291). This suggests a preference among Swiss voters for optimizing, rather than maximizing, the allocation of public resources for welfare programs (Roosma et al., 2013). This boils down to emphasizing the *security* goal of the welfare state rather than its *equality* goal, and it implies a commitment to the welfare state's central function of protecting people from the hardships of old age, extreme poverty, sickness and disability, and unemployment. In contrast, equality-seeking policies in the domains of housing, child care, or employment (e.g., family benefits, active labor-market policies) may be viewed as less vital for income maintenance and redistribution, and hence are less popular among Swiss citizens. In particular, state protection against job discrimination (one of our test items below) may seem too remote from the core functions of the welfare state to receive significant support.

In short, Swiss citizens are accustomed to seeing debates on welfare state reform as a routine exercise in trade-off and compromise, rather than as a deep conflict between irreconcilable cultures of the state. Also consistent with the assumed legitimacy of retrenchment arguments is the absence of significant bias in media coverage of voting issues. As far as newspaper reporting is concerned, there is no particular slant in the presentation of issues, which means that pro-welfare and anti-welfare opinions are given about the same chance of influencing public debates on welfare reform in Switzerland (Marquis et al., 2011).

4.2. Measurement strategy

SDR has been measured by a variety of techniques (for reviews, see Nederhof, 1985; King and Bruner, 2000; Paulhus, 2002; Tourangeau and Yan, 2007; Krosnick and Presser, 2010). Astute interviewing procedures have been experimented to elicit unbiased opinions by manipulating the attractiveness of socially desirable responses. In methods such as "the bogus pipeline", respondents are warned that any lie or exaggeration will be detected by a "physiological monitoring device" and are thus encouraged to provide reliable answers (Jones and Sigall, 1971; Dwight and Donovan, 2003). But the same goal can be achieved by inducing respondents to believe exactly the opposite — that there is no way to know what they actually responded! One version of this technique is known as the "Item Count Technique" (Droitcour et al., 1991) or as the "List Experiment" (e.g., Sniderman and Carmines, 1997). As it is the method that we selected to assess SDR on welfare state issues, it deserves further explanation.

The *List Experiment* has been used in various settings involving sensitive social issues.[10] For example, in the domain of racial politics, the basic purpose of the technique "is to persuade respondents that they can express hostility toward blacks without anyone's being aware that they have done so" (Kuklinski et al., 1997: 327).

[10] These issues include voting for a female or Black US President (Streb et al., 2008; Flavin and Keane, 2008; Heerwig and McCabe, 2009), support for a Jewish presidential candidate (Kane et al., 2004), attitudes toward immigration (Janus, 2007, 2010) and same-sex marriage (Lax et al., 2014), allowing illiterate people to vote in Lebanon (Corstange, 2009), and support for liberal racial policies and assistance for minorities (Kuklinski et al., 1997; Sniderman and Carmines, 1997; Sniderman and Hagendoorn, 2007; Zigerell, 2011).

We set up a similar experiment to gauge the degree of "conformism" in opinions about the welfare state. We chose to focus on issues which are potentially divisive but may enjoy some "fake consensus" because of social desirability. These issues were summarized in three statements on which subjects were invited to express their opinions:

1. *"Foreigners should have the same rights to social services as Swiss citizens, if they have lived and worked for a long time in Switzerland"* **(welfare entitlement)**
2. *"Foreign job seekers with equal skills and experience should not be discriminated against in favor of Swiss job seekers"* **(job nondiscrimination)**
3. *"Taxes should be collected more fairly, even if I or my family were to be taxed more heavily"* **(tax fairness)**

Statements 1 and 2 were conceived to appeal to the value of *equality*, but in such a way as to make an alternative frame, namely *"welfare chauvinism"*, immediately relevant and applicable. Besides, the concreteness of these two issues was low (Statement 1) to moderate (Statement 2), and the wording of questions was impersonal. As for Statement 3, it was conceived to prime the value of *fairness* (or equity) while also activating self-interested considerations. In contrast to Statements 1 and 2, it was quite concrete in its practical implications and was termed in personal terms ("I or my family"). As formulated in our Research Question, it is the purpose of the forthcoming empirical analysis to assess whether these variations in item features are associated with differences in SDR level.

All these statements constitute **"test items"**, as we call them, whose goal is to present a socially acceptable position that most students probably share *or feel compelled to agree with*. This expectation was largely met: Only 6 percent, 18 percent, and 12 percent of subjects expressed a "socially reprehensible opinion" by saying that they disagreed with the test items 1, 2, and 3, respectively. In other words, a very large majority of subjects concurred with the prevailing view in the students' community.

Each of the test items was embedded in a **list of four items** along with three **"baseline items"**, i.e., three other statements which a pre-test had shown to have well-defined response expectancies (see Appendix A for the list of all statements). Two of these additional statements may be considered "universally acceptable" and the third one "universally unacceptable".[11] The purpose of the baseline items is to determine whether, and how far, the large consensus on test items is actually the product of social desirability. As a matter of fact, the experimental design relies on the fact that *only some participants* were required to express *overt* opinions on the given statements. The subjects were randomly divided into three groups: a control treatment (hereafter CTRL), a first test treatment (ITEM4), and a second test treatment (ITEM3). These three groups are roughly similar in size (N=71, 66, and 68, respectively) and are identical from the perspective of all variables to be used in the empirical analysis (all *ps* from χ^2 tests > .57).[12]

In the **CTRL** condition, subjects were asked to give (overt) opinions about each of the 12 Statements (i.e., 3 Lists × 4 Statements). This group constitutes a baseline

[11] For two thirds of items, the ratio of unexpected responses (i.e., responses that contradicted the a priori acceptability or unacceptability criteria) did not exceed 7% of valid responses.

[12] One significant exception is the fact that there are twice as many binationals and foreigners in the ITEM3 group as in each of the two other groups (χ^2=6.70, p=0.035). Given the random assignment to groups, the concentration of half the number of binationals and foreigners in the ITEM3 group can only be explained, to the best of our knowledge, by the kind of "randomization flaws" that *must* occur every now and then with low-N samples.

condition against which to gauge the amount of social desirability.[13] It is the basis for calculating the acceptance ratio of the test item under conditions where SDR is expected to occur (overt expression of opinions).[14] In the **ITEM4** condition, subjects were *not* asked to offer opinions on the statements. Instead, the question read: "Below you will find a list of arguments that are often voiced in public debates. After reading all statements, please indicate *how many* statements you (rather or definitely) agree with. We do not want to know which statements you would approve, but only how many — that is, from 0 to 4". In the **ITEM3** condition, subjects were presented with the same list of statements, but this time *without the test item*, and asked to indicate the number of statements they approved (i.e., from 0 to 3).

In answering such a question, subjects in both ITEM4 and ITEM3 conditions probably realized that there is no way for the experimenter to determine whether the test item is part of the items they refuse, nor to determine which items are refused or accepted whatsoever (unless they accept or refuse all items, of course). Indeed, it is impossible to know which subjects accepted or refused the test item. But the experimenter can determine the *aggregate* proportion of subjects who answered either way. To do so, she just needs to compare the figures that she gets from the two test conditions. In other words, if we assume the successful randomization of subjects in the various conditions, subtracting the average number of accepted items in the ITEM3 condition from the average number in the ITEM4 condition yields a rough estimate of the *true acceptance ratio for the test item*. This estimate is supposedly purged of social desirability bias, in contrast to the acceptance ratio measured through direct questioning in the CTRL group. Accordingly, the extent of social desirability responding on the test item can be easily computed as a difference-in-means estimator: **SDR = $M_{CTRL} - (M_{ITEM4} - M_{ITEM3})$**. In the example below, one can estimate that 16 percent of subjects probably lied when agreeing with the test item in the CTRL condition:

CTRL			**ITEM4**	**ITEM3**
Indicate whether you agree with the following statements	Yes	No	*Indicate* how many *statements you agree with*	*Indicate* how many *statements you agree with*
• Baseline item 1	☐	☐	• Baseline item 1	• Baseline item 1
• Baseline item 2	☐	☐	• Baseline item 2	• Baseline item 2
• Test item	☐	☐	• Test item	• Baseline item 3
• Baseline item 3	☐	☐	• Baseline item 3	
			I agree with ___ items	I agree with ___ items
M_{CTRL} = 0.85 (85% agreeing w. test item)			M_{ITEM4} = 2.92 items	M_{ITEM3} = 2.23 items

SDR = $M_{CTRL} - (M_{ITEM4} - M_{ITEM3})$ = 0.85 – (2.92 – 2.23) = 0.16

This difference-in-means method has been applied since the invention of the List Experiment technique in the early 1990s. Comparing counts *within specific subgroups* (e.g., subjects with low vs. high education) enabled researchers to identify plausible predictors of SDR. This method, however, hardly allows for a true "multivariate"

[13] This CTRL condition does not appear as such in the seminal studies that introduced the List Experiment procedure into political science (e.g., Kuklinski et al., 1997; Sniderman and Carmines, 1997). Their CTRL (i.e., "baseline") condition corresponds to the ITEM3 group here, while their test condition is the present ITEM4 group. In fact, the designations used in the various studies are only a matter of definition.

[14] It may be noted that subjects in the CTRL group are asked to indicate their approval of items on four-grade scales ranging from "definitely disagree" to "definitely agree", with "rather disagree" and "rather agree" as intermediate response categories. Unlike with dichotomous response choices (such as in the example above), the possibility to express a "mild" opposition to the test item may elicit more honest responses and therefore reduce SDR.

analysis.[15] Fortunately, considerable progress has been made in recent years in the development of a true multivariate framework for analyzing list experiment data. In a first decisive step, Glynn (2010, 2013) and Imai (2011) have drawn attention to hitherto unexplored properties of count lists. They have shown that it is possible to determine the *joint* distribution of approval of the test item and approval of baseline items. For example, one can compute the aggregate proportion of respondents that honestly approve of the test item *and* approve of exactly two baseline items. To demonstrate this, Table 1 takes the example of our List 2 (containing the sensitive item "job nondiscrimination"). It first displays the relative frequencies of the number *k* of approved items among the ITEM4 and ITEM3 groups (rows 1 and 3), as well as the cumulative frequencies of approving *at least k* items (rows 2 and 4). Because the only difference between the ITEM4 and ITEM3 lists is the sensitive item, one can subtract the cumulative frequencies and deduce the estimated proportion of respondents in the population who *jointly* endorse the sensitive item and any *k*-1 number of baseline items. As Glynn (2010: 13) observes, the sum of all these joint proportions (0.713) is strictly equal to the difference-in-means estimate (2.344 – 1.631 = 0.713).

In several respects, however, the joint proportions from Table 1 are much more informative than the difference-in-means estimate. Going one step further, one can estimate the probability of approving the test item, conditional on the number of accepted items. For example, we know from row 5 in Table 1 that an estimated 31.9% of population respondents accepted the test item and two control items [2,1], that is, a total of three items. In turn, this proportion can be compared to the 34.4% of ITEM4 respondents who accepted three items, no matter which type (i.e., either three control items [3,0] or two control items and the test item [2,1]). The ratio of these two proportions, 0.319/0.344, indicates that an estimated 92.5% of respondents from the ITEM4 group who reported agreeing with three items endorsed the test item as part of these three accepted items (figures for this example are **in bold** in Table 1). Likewise, the same 31.9% estimate of the population who accepted the test item and two baseline items can be compared to the 44.6% of ITEM3 respondents who accepted two items (and were not given the opportunity of choosing the test item anyway); the resulting ratio of 0.319/0.446 indicates that an estimated 71.4% of respondents from the ITEM3 group who reported agreeing with two items *would have* accepted the test item if given the chance to do so (figures for this example are underlined in Table 1).

Table 1: Estimated joint and conditional probabilities from List 2

	Number of reported items (*k*)					
	0	1	2	3	4	Sum
1. Proportion reporting exactly *k* items (ITEM4)	0.033	0.098	0.443	**0.344**	0.082	1.000
2. Proportion reporting at least *k* items (ITEM4)	1.000	0.967	0.869	0.426	0.082	
3. Proportion reporting exactly *k* items (ITEM3)	0.031	0.415	0.446	0.108	0.000	1.000
4. Proportion reporting at least *k* items (ITEM3)	1.000	0.969	0.554	0.108	0.000	
5. Estimated population *P* endorsing the test item **and** *k*-1 control items (row 2 – row 4)	0.000	0.000[a]	0.315	**0.319**	0.082	0.713
6. Estimated conditional probability (ITEM4)	0.000	0.000	0.712	**0.925**	1.000	
7. Estimated conditional probability (ITEM3)	0.000	0.758	0.714	0.761	—	

Note: [a]: the real value (-0.002) was truncated to zero, because it is nonsensical and most likely due to small sample size (see Glynn 2010: 14-19). This problem is alleviated in the model presented below.

[15] In trying to control for multicollinearity and omitted variable bias, one can split the data and run difference-in-means tests in separate subsamples. But breaking down subjects into increasingly small subgroups leads to degrees-of-freedom problems (e.g., less precise and less statistically significant estimates) without really alleviating the above problems.

As our last sentence above suggests, the procedure used in Table 1 depends on the effective randomization of subjects between the ITEM3 and ITEM4 experimental conditions. Of course, this implicit assumption will tend to be more easily verified as the sample size increases. Nonetheless, the last rows of Table 1 can be considered the "fundamental building blocks for treating the sensitive item as missing data (e.g. multiple imputation, the EM algorithm, or data augmentation), which allows multivariate modeling of the sensitive item" (Glynn, 2010: 14). Indeed, the absence of a direct question about the sensitive item compares to situations where the variable of interest is missing. More importantly, however, the joint and conditional probabilities estimated in Table 1 provide considerable leeway for developing a true multivariate analysis of the determinants of the acceptance of the test item. Building on the properties of list experiments discussed above, Blair and Imai (2012; Blair et al., 2014) have developed an expectation-maximization algorithm whose aim is to model the impact of several individual characteristics on the joint distribution of test and baseline items approval.[16] Implemented through the R "List" package designed by Blair and Imai (2014), this multivariate regression analysis enables us to compute the predicted probability of "truly" accepting the test item in all subgroups relevant to the test of our hypotheses (see section 4.4). We denote this probability as **LISTT**. Likewise, Blair and Imai's procedure estimates the effects of the same set of predictors on the number of accepted baseline items (hereafter **LISTB**). In addition, we can compute the corresponding group-based probability of accepting the test item in the CTRL group via logistic regression, which we denote as **ACT**. The extent of SDR is then quite simply calculated, in each group, as **SDR = ACT – LISTT**.

This measurement strategy is predicated on two further assumptions — beyond that of experimental randomization. First, we have to assume that opinions about the test item, *and not about other items*, are responsible for the between-treatment difference in the number of accepted items. In other words, list experiments have to make the "***no design effect assumption***" that "the addition of the sensitive item does not change the sum of affirmative answers to the control items" (Blair and Imai, 2012: 51; see also Imai, 2011: 408-409). The R "List" package (Blair and Imai 2014) provides us with a statistical test to rule out the alternative hypothesis that answers to baseline items are indeed affected by the inclusion of the test item. Interestingly, however, the "no design effect" assumption does *not* require subjects to give honest responses to baseline items (either in direct or list count ways). SDR on baseline items is an independent question, which we will address in the concluding part of this article, because it may shed light on basic features of SDR. To this aim, the **LISTB** estimates (i.e., the group-based numbers of baseline items predictably accepted in the treatment conditions) can be compared to the corresponding amounts of baseline items predictably accepted by subjects in the CTRL condition (hereafter **ACB**, i.e., a "pseudo item count response" according to Tsuchiya et al., 2007). The latter estimates can be easily obtained via OLS regression, and SDR on baseline items is accordingly computed as **SDRb = ACB – LISTB**.

The second assumption is that respondents *do not lie* when reporting the number of items they agree with. Overall this "***no liar assumption***" is tenable, because list experiments were precisely designed to remove pressures from the social context and to lead respondents into believing that their true feelings are hidden to the experimenter. To some extent they are. As Blair and Imai (2012: 65-55) make clear, however, there are two situations in which privacy may be perceived to be threatened and, accordingly, responses may still be biased by social desirablity concerns. First, there is the possibility of a "ceiling effect", which "is caused by the

[16] Corstange (2009) has proposed a similar method based on a different experimental design. However, Blair and Imai (2012) argue that their method is more precise and reliable, especially when the sample size is small.

fact that privacy is not protected for those respondents in the treatment group whose true preferences are *affirmative for all the sensitive and control items*" (2012: 65; emphasis added). In this case, respondents know that their answers to each item are completely identified and may thus be tempted to decrease their item count by one unit to restore ambiguity. Second, there is the possiblity of a "floor effect", in which "some of the respondents whose truthful answer is affirmative only for the sensitive item (and thus negative for all control items) give [0] as an answer instead of the truthful [1]. This may occur when the control items are expected to generate many negative answers. In such a situation, the respondents in the treatment group whose truthful answer is affirmative only for the sensitive item may fear that their true preference for the sensitive item would be revealed by giving [1] as their answer" (2012: 65-66). It is important to note that both situations entail an artificial decrease in the average number of reported items in the treatment group. Fortunately, the R "List" package offers the possibility to estimate the extent of deflation in item counts stemming from ceiling and floor effects. Because the algorithm has to estimate two further unknown quantities, i.e., the proportion of ceiling and floor "liars", it has to introduce additional constraints. These constraints require auxiliary information from the covariates entered into the model to account for variations in the acceptance of the test item (see Blair and Imai, 2012: 65-68). Thus, the use of a multivariate setup can be seen both as an enlargement of the predictive value of list experiments and as a requirement to establish their validity against potential failures.[17]

4.3. Subjects

In total, 185 students from various disciplines at the University of Bern, Switzerland, were recruited to participate in the study, through voluntary participation during a political science lecture, through announcements in other lectures, and through posters and leaflets in halls. Most subjects came from curricula in political science (28%), economics (20%), and history (13%), but significant numbers of subjects were drawn from other fields such as literature and linguistics, geography, sociology, psychology, and other human sciences. In addition, 20 apprentices (masons and sanitary technicians) were enrolled at a training school in the same town.[18] Overall, more male subjects (63%) than female subjects (37%) participated in the study; likewise, the number of Swiss citizens (80%) greatly exceeded that of foreigners (20%). As regards age, a majority of participants were 22 or younger, but all ages between 17 and 32 years were represented in the sample (M=23.1, SD=5.7). Subjects recruited outside the political science lecture were paid 20 Swiss francs (about US$16 at the time) for their participation.

The choice of students as experimental subjects is both pragmatic and theory-driven. Leftist orientations of students have been proverbial since the campus unrest of the late 1960s in the United States, France, or Germany. As a matter of fact, most studies of students' values and attitudes have uncovered evidence that the student

[17] For further discussion of the "no design effect" and "no liar" assumptions, see Zigerell (2011) and de Jonge and Nickerson (2012). In particular, Zigerell (2011) considers the possibility (and finds some evidence) that, in situations where ceiling and floor effects are expected, item counts may be deflated (and sometimes inflated) by more than one unit. These most extreme violations of the "no liar" assumptions might occur when respondents are particularly eager to dissociate from strongly undesirable behaviors or traits, or to associate with strongly desirable ones.

[18] The rationale for recruiting apprentices was to increase the variance in key variables, such as chronic and issue-specific knowledge, issue involvement, age, and ideology. However, because apprentices are arguably a different population from that of students, we controlled for the specific contribution of the apprentice data to our SDR estimates. None of the conclusions of our empirical analysis were substantially altered by including or excluding apprentices.

"subculture" is biased toward liberal/leftist views, even though this liberal advantage has ebbed and flowed in recent decades (e.g., Dey, 1997; Pryor et al., 2012). Importantly, the distribution of political preferences varies between academic fields. These differences reflect either *socialization*, whereby faculty and peer group influence within each academic field tends to make students' preferences more homogenous over time (e.g., Alwin et al., 1991; Liu et al., 2009), or *self-selection*, whereby initial preferences determine which academic subjects are chosen in the first place and no further "indoctrination" is to be expected from curriculum experiences (e.g., Kemmelmeier et al., 2005; Hastie, 2007). Most probably, both mechanisms are at work, but in varying degrees according to the type of academic discipline.[19]

This being said, normative influence also operates in a larger sociopolitical context than mere curriculum or faculty (e.g., Dey, 1997; Pascarella and Terenzini, 2005: 293-295; Hurtado et al., 2012). As already noted, "campus norms" are extremely important and may overwhelm the influence of other peer groups. However, the influence of population attitudes (at the city, regional, or national level) should not be underestimated. It is perhaps worth noting that the University of Bern, where our experiment took place, is situated in the German-speaking area of Switzerland. Compared with their French- and Italian-speaking fellow citizens, Swiss Germans have been more reluctant to develop welfare state institutions and to defend them against retrenchment.[20]

4.4. Moderator variables

As we argued in the presentation of our hypotheses, SDR is conditional on a number of situational and personal characteristics, which may either facilitate or inhibit the effect of social desirability pressures. Ten potential moderator variables were selected for their applicability in the context of this study: general and issue-specific political knowledge, left-right self-placement, age, sex, academic discipline, belief activation, personal salience of welfare state issues, daily time devoted to political information, and nationality (Swiss vs. foreigner or binational). Some of these moderators are expected to be involved both in the "private" evaluation of items and in the response editing process leading to their "public" acceptation. Hence, the scaling of variables permitting, nonlinear relationships with SDR are conceivable. The construction of moderator variables is described in Appendix B.

---

[19] For example, political attitudes of students (and faculty staff) in *economics* are commonly biased in a right-wing direction. This tendency is more likely to stem from *self-selection* in the economics curriculum than from indoctrination through economic education (Carter and Irons, 1991; Manley et al., 2001; Kemmelmeier et al., 2005). Economists tend to attach more importance to self-enhancement values and less importance to universalistic values and electoral participation (Gray and Wuffle, 2005; Gandal et al., 2005). Likewise, they tend to behave in a more selfish way, for example acting as "free riders" in the provision of public goods (Frey and Meier, 2003). Indeed, a similar right-wing bias obtains in our student sample, where the economists' mean left-right self-placement of 4.30 (on a 0-10 scale) is statistically different from the political scientists' mean of 3.24 ($t=3.37$, $p<.002$) and the historians' mean of 2.46 ($t=4.91$, $p<.001$). By usual standards, however, a majority of economics students in our sample would qualify as left-wingers. To be sure, this may indicate that social desirability already occurs in the ideological self-description of some students.

[20] German-speaking citizens tend to put more emphasis on individual responsibility and work ethics than Latin citizens. This is immediately visible in the between-region differences in voting patterns on welfare state issues (Obinger et al., 2005: 341; Freitag and Vatter, 2006: 100). Survey evidence (Eugster et al., 2011) shows that German-speaking citizens' support for welfare policies is comparable to the support measured in liberal countries (e.g., US, Australia, Japan), whereas French-speaking Swiss are much closer to the norm of less market-oriented neighbour countries such as Italy and France.

The "belief activation" variable deserves further consideration. A procedure was introduced to manipulate the *number of accessible beliefs about welfare state issues* respondents have in mind when undertaking the list experiment task. This procedure is directly inspired from the ANES "likes and dislikes" questions about parties and presidential candidates. Prior to the SDR questions, subjects were asked to list the first five reasons that come to mind for supporting *and* opposing the welfare state and Swiss EU membership, respectively.[21] The real purpose of the questions was to conjure up *contradictory* "ideas" and feelings about the welfare state, including social normative reasons subjects may not really endorse but feel convenient to express at the beginning of the experiment. In sum, by asking subjects to think in detail about welfare state issues, the procedure was meant to "refresh" and diversify their relevant beliefs, to make them more accessible in memory, and thus to foster *ambivalence* on welfare state issues. According to Hypothesis 1, this should lead to heightened SDR.

Importantly, the belief activation questions were posed to only half the subjects in each of the CTRL, ITEM4, and ITEM3 conditions. Belief activation thus constitutes a second factor, orthogonal to the variation in the format of SDR questions derived from the List Experiment (see Table 2). For purposes of analysis, though, belief activation will be used as a *moderator variable* in SDR, much in the same way as subjects' personal characteristics such as political knowledge, academic training, or age.

Table 2: The experiment's 2×3 factorial design, with number of subjects in parentheses

Belief activation	**Type of SDR questions**			
	CTRL	ITEM4	ITEM3	Total
Non-activation	(36)	(36)	(36)	(108)
Activation	(35)	(30)	(32)	(97)
Total	(71)	(66)	(68)	(205)

[21] The questions read: "Why is the welfare state a good idea? What are the positive consequences? Please write down the first ideas that come to mind! (…) In contrast, what is it about the welfare state that you do not find so appealing? Please write down the first ideas that come to mind!" Similar questions were posed about the "idea of joining the EU". These pros and cons questions were not meant as a memory probe (subjects were instructed not to try "to fill the whole list at any price") nor even as a measure of overall attitudes toward the two issues.

5. Empirical results

5.1. Assessing the covert and overt acceptance of test items

Multivariate analysis was applied to our experimental data to estimate the effects of individual and situational variables on the various components of item agreement:

- **LISTT and LITSB**: Blair and Imai's ICTREG algorithm (R "List" package; Blair and Imai 2014) was used to estimate effects on item counts in list conditions;
- **ACT and ACB**: logistic and linear regression was used to estimate effects on agreement with the test and baseline items in the CTRL condition.

The results of these analyses are presented in Tables 3 to 5. Before we comment on the results, it is necessary to explain how the ACT model was built. As can be seen from the tables, the effects of some variables are left aside. To explain why this simplification of our models was unavoidable, it should be stressed that the full models fitted the data very poorly (in all lists) and that estimated coefficients were, with few exceptions, non-significant and obviously unreliable. Thus the poor fit of the ACT models reflected real problems in estimating the effects of the various predictors. First, with as many as 14 predictors and little more than 60 observations, degrees of freedom proved insufficient to allow for reliable estimates. A second problem was that, as stressed above, our dependent variables are highly skewed toward agreement with the test items (94%, 82%, and 88% for Lists 1, 2, and 3, respectively). Given this distribution, the presence of "quasi-separation" usually leads to systematic biases in estimated coefficients when conventional methods are used.[22] As a matter of fact, this problem proved intractable in List 1, where all logistic regression coefficients and standard errors were immensely inflated and therefore useless for estimation purposes, but the problem also loomed large in Lists 2 and 3.

When confronted with a degrees-of-freedom and/or a quasi-separation problem, classical regression models will tend to produce unreliable coefficients, and thus alternative estimation methods are needed. Two strategies were used here to help circumvent these problems. First, *penalized maximum likelihood (PML)* was chosen as an alternative to traditional maximum likelihood; it is probably the best suited estimation method when data are completely or nearly separated (Heinze, 2006). Second, we used backward variable selection (based on likelihood ratio tests) to remove "noisy" variables whose inclusion in the model prevents a better estimation of "real" predictors.[23]

Before turning to the interpretation of the LISTT models in Tables 3 to 5, we must consider the possibility that the underlying assumptions of these models are violated

[22] Quasi separation, i.e., the fact that a linear combination of independent variables predicts the dichotomous outcome almost perfectly, is all the more likely and problematic as the sample size is small and the number of variables in the model is large (Hosmer and Lemeshow, 2000: 139).

[23] We put two constraints on the backward selection process. First, we stopped the procedure as soon as at least two predictors were significant ($p<.10$). Second, to enhance our capacity to compare effects between treatments and thus to ascertain SDR, we made sure that all variables which appeared as significant predictors ($p<.05$) in the LISTT regression were included in our models. The only exception was the exclusion of the "foreigner" variable from the ACT model for List 3. The effect of this variable was still impossible to determine even using PML analysis, because estimates were obviously unreliable in all tested models. As a matter of fact, the huge coefficients and standard errors point to the fact that our data are highly separated with respect to the "foreigner" category: all subjects rejecting the "tax fairness" item were Swiss citizens, and all foreign subjects approved of it.

(see section 4.2). First, we must rule out the possibility that the inclusion of the test item in the long list did not affect respondents' answers to baseline items. Our examination of this "no design effect" assumption provided reassuring results. None of the Bonferroni-corrected significance tests indicates that answers to baseline items might have been "contaminated" by adding the test item.[24] Second, we must examine the "no liar" assumption that respondents' answers to the count questions are indeed truthful. As explained above, the existence of ceiling and floor effects would question this assumption, because such effects imply that some respondents deliberately misreport the total number of items they agree with in order to avoid being associated with the test item. The test of this assumption, using the algorithm available in the R "List" package, could not be performed in its full form (i.e., allowing an estimation of each covariate's contribution to the amount of ceiling and floor lying), because degrees of freedom were insufficient and problems of complete separation (also due to small sample size) led to computational problems.[25] Instead, we ran an intercept-only model (assuming that the propensity to "lying" was roughly the same in all student subgroups) to estimate the population proportion of ceiling and floor "liars" for each List. This test yielded a quite moderate proportion of floor liars, varying between 0.4% (List 3), 1.0% (List 1), and 2.4% (List 2). Likewise, the proportion of ceiling liars was modest in List 2 (0.8%) and List 3 (3.9%); however, it was much higher in List 1 (15.7%).[26]

Overall, then, violations of the "no liar" assumption were limited in scope, and do not appear to threaten the validity of the whole procedure *except in List 1*. Based on these tests of formal assumptions, we might stop here our analysis of List 1 and proceed with the other lists. Indeed, we will avoid drawing any independent conclusion from List 1 results. However, we decided to maintain this list in our forthcoming multivariate analysis to see whether corroborating evidence may be found for the role of moderators in SDR. To the extent that moderators play similar roles in List 1 and in the other two lists, this would tend to reinforce our confidence in our results.[27] The other way round, we will refrain from interpreting findings from List 1 when they lead to different conclusions than findings from the other lists.

Looking at Tables 3 to 5, there appears to be remarkably few similarities across Lists; that is, there are few (if any) effects which persist from one set of items to the next. First, general knowledge is systematically related to the acceptance of the test items (though the relationship is only marginally significant in List 3). The more

[24] All Bonferroni-corrected p-values from our three Lists are greater than .30, leading us to accept the null hypothesis of no design effect.

[25] The problem of (quasi) complete separation is not uncommon in list experiments. However, the solution suggested by Blair and Imai (2012: 67), namely to add "weakly informative priors" to the model for floor/ceiling effects (implemented through the *bayesglm* tool in the R "List" package), does not work for our data — probably because *several* variables are involved in separation. On the one hand, our model may indeed be too complex and unstable; Graeme Blair (personal communication, August 2016) acknowledges that the number of variables in the model "is likely one problem" because "these models are fragile and hard to estimate". On the other hand, the common solution of omitting the best predictors to prevent separation is suboptimal and leads to specification bias (Zorn 2005: 161-162; Gelman et al. 2008: 1361); in addition, the comparability and sheer utility of the model is threatened if its specification changes from one list to the next and does not correspond to the multivariate moderator model tested below. Accordingly, an intercept-only model was tested to estimate *overall* proportions of floor and ceiling liars.

[26] Due to the computational problems mentioned above, no standard error could be estimated for the population proportions of ceiling and floor liars. However, it seems fair to say that only List 1 produces estimates which are undoubtedly biased by floor and ceiling lying behavior.

[27] Although the results for List 1 are made unreliable by the estimated 15 percent of "floor liars", the likelihood that a moderator would show a reverse effect in the absence of lying is extremely small.

subjects knew about Swiss political institutions and leaders, the more likely they were to accept the sensitive statements about welfare entitlement, job nondiscrimination, and tax fairness. Second, age also tends to be associated with test item acceptance, as younger subjects were less favourable toward the items than were older subjects — but the effect is only significant in List 1, which is plagued by ceiling effects (see above). Third, personal salience is also related to the acceptance of test items in two Lists. In this case, however, the effect is inconsistent: subjects who expressed high concern about welfare state issues were more likely to accept the tax fairness item (List 3) but *less* likely to accept the job nondiscrimination item (List 2). Other effects emerged in some Lists (such as the reduced acceptance of the "tax fairness" item among foreigners / binationals and political science students in List 3) but failed to show up in at least another List.

Likewise, the ACT models did not show a discernible pattern of effects on the *overt* acceptance of test items. Yet a couple of effects may be noted for single Lists. Middle-aged subjects (22-23 years olds) were less likely to support the tax fairness item than their older counterparts. Similarly, younger subjects (up to 21 years) were less likely than their elders to endorse the job nondiscrimination item. In contrast, this item was largely supported by subjects with longer exposure to political news. Finally, it may be noted that extreme left subjects were more likely than center-right subjects to accept the tax fairness statement.

One reason for the lack of clear-cut results is related to the difficulties to predict the endorsement of statements which — at least in *public* situations — are virtually universally accepted, so that little variation is left to be explained. To account for the (few) differences that did exist in the answers of the various groups, one may argue that certain categories of subjects were simply better aware of where the social norm stood, while others were not (or misperceived the norm). Accordingly, it may be that certain categories (e.g., older, better informed subjects) knew how to adjust their answers to the norm, while others did not. *This is precisely what SDR analysis is all about*. Therefore, we should not take the scattered results from our ACT models as evidence that SDR did not occur in our experiment. It is only through *comparing overt and covert answers* that the underlying factors and mechanisms of SDR may be uncovered — a task to which we turn in the next section.

Table 3: Multivariate analysis of the List Experiment questions and the direct questions (List 1; test item = "welfare entitlement")

	Test item, list estimates (LISTT)			Baseline items, list estimates (LISTB)			Test item, direct question (ACT)			Baseline items, direct questions (ACB)		
	Est.	S.E.	p	Est.	S.E.	p	Est.	S.E.	p	Est.	S.E.	p
Intercept	1.222***	0.386	0.002	1.793***	0.274	0.000	1.401+	1.188	0.238	1.869***	0.181	0.000
General Knowledge = high	0.463**	0.228	0.045	-0.024	0.080	0.766	1.052	1.328	0.428	-0.057	0.102	0.582
Specific Knowledge = high	0.048	0.198	0.811	-0.016	0.094	0.866	1.781+	1.407	0.206	0.106	0.097	0.278
Daily Information > 30 min.	0.167	0.253	0.509	-0.172+	0.119	0.151	-0.749	1.379	0.587	0.088	0.112	0.433
Personal Salience = high	-0.070	0.228	0.761	0.192+	0.136	0.161	0.858	1.092	0.432	-0.034	0.088	0.702
Age = up to 21 y.	-0.638**	0.277	0.024	0.134+	0.109	0.222	1.794+	1.363	0.188	0.072	0.120	0.553
Age = 22-23 y.	-0.210	0.199	0.293	0.104+	0.087	0.234	2.643*	1.596	0.098	0.016	0.115	0.889
Sex = Female	0.024	0.231	0.918	-0.048	0.104	0.644				-0.107	0.111	0.340
Ideology = Extreme Left	-0.100	0.235	0.670	0.060	0.108	0.581	-1.309	1.528	0.392	0.065	0.127	0.612
Ideology = Center / Right	-0.389+	0.321	0.228	-0.101	0.179	0.574	-1.175	1.294	0.363	-0.181+	0.120	0.139
Discipline : Political Science	-0.443+	0.281	0.118	0.179	0.174	0.305				0.064	0.149	0.670
Discipline : History	-0.269	0.388	0.490	-0.059	0.229	0.799				-0.050	0.176	0.777
Discipline : Others	-0.168	0.335	0.617	-0.029	0.157	0.855	-2.334*	1.309	0.075	0.054	0.145	0.713
Nationality = Foreigner / Binat.	0.182	0.229	0.428	-0.033	0.127	0.797				0.075	0.162	0.646
Activation of Beliefs = Yes	-0.128	0.163	0.435	0.072	0.085	0.396	1.171	1.120	0.296	-0.037	0.087	0.674

Notes: ***: p<.01; **: p<.05; *: p<.10; +: p<.25. The reference categories for the various predictors are as follows: General Knowledge = low; Specific Knowledge = low; Daily Information < 30 min.; Personal Salience = low; Age = 24 years and older; Sex = Male; Ideology = Moderate Left; Discipline = Economics; Nationality = Swiss; Activation of Beliefs = No. For the ACT model, reference categories include those modalities for which no regression coefficient could be computed (e.g., the reference for academic discipline includes economics as well as political science and history).

Table 4: Multivariate analysis of the List Experiment questions and the direct questions (List 2; test item = "job nondiscrimination")

	Test item, list estimates (LISTT)			Baseline items, list estimates (LISTB)			Test item, direct question (ACT)			Baseline items, direct questions (ACB)		
	Est.	S.E.	p	Est.	S.E.	p	Est.	S.E.	p	Est.	S.E.	p
Intercept	-0.609	0.622	0.330	2.490***	0.490	0.000	1.975*	1.188	0.096	2.004***	0.332	0.000
General Knowledge = high	0.839**	0.351	0.019	-0.451*	0.248	0.072	-0.716	0.946	0.449	0.015	0.180	0.935
Specific Knowledge = high	0.602**	0.303	0.050	-0.514**	0.212	0.017	0.832	0.906	0.359	0.195	0.183	0.293
Daily Information > 30 min.	0.694+	0.431	0.111	-0.318+	0.252	0.211	2.170**	1.027	0.035	-0.416**	0.197	0.042
Personal Salience = high	-1.165***	0.421	0.007	0.426+	0.260	0.105	0.048	0.713	0.947	-0.102	0.160	0.529
Age = up to 21 y.	-0.255	0.370	0.492	0.059	0.229	0.799	-1.832*	1.078	0.089	-0.008	0.216	0.971
Age = 22-23 y.	-0.502+	0.371	0.178	0.079	0.281	0.778	-0.881	1.025	0.390	-0.404*	0.213	0.066
Sex = Female	0.353	0.357	0.325	-0.133	0.223	0.553				-0.509**	0.197	0.014
Ideology = Extreme Left	0.325	0.443	0.465	-0.460+	0.283	0.107				0.321+	0.244	0.197
Ideology = Center / Right	0.479	0.468	0.309	-0.159	0.277	0.568	1.012	1.014	0.318	-0.259	0.223	0.253
Discipline : Political Science	0.388	0.548	0.480	-0.024	0.419	0.955				0.421+	0.259	0.112
Discipline : History	-0.100	0.907	0.913	0.106	0.452	0.814				0.120	0.335	0.723
Discipline : Others	0.864*	0.510	0.094	-0.514+	0.384	0.184	-0.690	0.760	0.364	0.552**	0.268	0.047
Nationality = Foreigner / Binat.	0.506+	0.347	0.148	-0.564**	0.258	0.031				0.174	0.283	0.544
Activation of Beliefs = Yes	0.554*	0.310	0.077	0.002	0.215	0.991	-0.338	0.713	0.636	-0.031	0.156	0.843

Notes: See Table 3.

Table 5: Multivariate analysis of the List Experiment questions and the direct questions (List 3; test item = "tax fairness")

	Test item, list estimates (LISTT)			Baseline items, list estimates (LISTB)			Test item, direct question (ACT)			Baseline items, direct questions (ACB)		
	Est.	S.E.	p	Est.	S.E.	p	Est.	S.E.	p	Est.	S.E.	p
Intercept	1.333**	0.532	0.014	2.067***	0.352	0.000	1.630+	1.271	0.200	2.399***	0.242	0.000
General Knowledge = high	0.398+	0.293	0.177	-0.175	0.175	0.322	-0.633	1.103	0.566	0.011	0.134	0.936
Specific Knowledge = high	0.052	0.268	0.846	-0.141	0.163	0.389	1.034	1.186	0.383	-0.188+	0.128	0.149
Daily Information > 30 min.	-0.109	0.321	0.736	0.059	0.182	0.749				-0.133	0.149	0.378
Personal Salience = high	0.556**	0.269	0.042	0.112	0.180	0.534	0.503	0.847	0.552	0.153+	0.118	0.202
Age = up to 21 y.	-0.514+	0.316	0.107	-0.099	0.200	0.623				0.059	0.160	0.712
Age = 22-23 y.	-0.242	0.327	0.462	-0.061	0.213	0.776	-1.920*	1.031	0.063	0.088	0.152	0.564
Sex = Female	0.092	0.307	0.765	-0.137	0.145	0.347	-0.950	0.896	0.289	0.038	0.145	0.796
Ideology = Extreme Left	-0.494+	0.327	0.135	-0.131	0.193	0.498	2.923*	1.740	0.093	0.273+	0.175	0.125
Ideology = Center / Right	-0.591+	0.420	0.163	-0.072	0.224	0.749	-0.896	1.248	0.472	0.356**	0.158	0.029
Discipline : Political Science	-0.705*	0.380	0.067	0.598**	0.261	0.024	0.246	1.393	0.860	-0.396**	0.195	0.048
Discipline : History	-0.694+	0.489	0.159	0.322	0.304	0.291	-0.609	1.387	0.661	-0.362+	0.234	0.129
Discipline : Others	-0.328	0.417	0.433	0.258	0.231	0.266	0.878	1.448	0.544	-0.291+	0.194	0.139
Nationality = Foreigner / Binat.	-0.752**	0.362	0.040	0.363*	0.187	0.055				-0.406*	0.206	0.055
Activation of Beliefs = Yes	0.170	0.289	0.558	-0.188	0.186	0.313	0.494	0.819	0.547	-0.088	0.114	0.444

Notes: See Table 3.

5.2. Assessing Socially Desirable Responding (SDR)

We begin to implement our measurement strategy with respect to the *overall* amount of SDR in each of the three Lists. In this case, estimating SDR is quite straightforward:[29] SDR = ACT – LISTT + LISTB, with standard error

$$\hat{\sigma}_{\text{SDR}} = \sqrt{\frac{s^2_{\text{ACT}}}{N_{\text{ACT}}} + \frac{s^2_{\text{LISTT}}}{N_{\text{LISTT}}} + \frac{s^2_{\text{LISTB}}}{N_{\text{LISTB}}}}\,.$$

A SDR score of zero indicates a strict equivalence of estimates for overt and covert treatments and hence the lack of any tendency to hide one's true preferences. Table 6 gives the overall estimates for each List. As it turns out, the amount of SDR is rather small, as it ranges between 8 percent (List 1) and 18 percent (List 3). (Recall, however, that results for List 1 are not reliable due to ceiling effects, which probably leads to an underestimation of SDR in this list). In addition, the estimate is (marginally) significant only in List 3 ($p<.07$, one-tailed test). Our evidence thus suggests that few people hide their true preferences on socially sensitive topics.

Table 6: Acceptation of items and estimates of SDR (all subjects)

	Acceptation of test item (ACT)			Number of accepted items (LISTB)			Number of accepted items (LISTT)			SDR estimate (ACT – LISTT + LISTB)		
	%	Std. dev.	N	Mean	Std. dev.	N	Mean	Std. dev.	N	%	s.e.	p (1-tailed)
LIST 1 (welfare entitl.)	94.3	23.4	70	1.912	0.334	68	2.773	0.627	66	8.2	9.2	0.186
LIST 2 (job nondiscr.)	81.7	39.0	71	1.632	0.710	68	2.318	0.880	66	13.1	14.6	0.185
LIST 3 (tax fairness)	88.1	32.7	67	2.162	0.536	68	2.864	0.742	66	17.9	11.9	0.068

Overall, SDR on the tax fairness issue is modest and not significantly larger than SDR on the welfare entitlement and job nondiscrimination issues. However, as a tentative answer to our Research Question (section 3.3), these differences suggest that SDR increases as a function of some "concreteness" gradient. Leaving the welfare entitlement issue aside, we note that the job nondiscrimination issue is only moderately concrete and personalized, as subjects might worry about a *future* labor market competition with foreigners. In contrast, the tax fairness issue elicits the greatest extent of SDR, perhaps because it promises to have the most concrete and immediate consequences for one's material situation. Although comparative evidence is generally lacking, this interpretation is consistent with findings showing little SDR on "abstract" and unobtrusive issues such as the environment (Milfont, 2009).

However, these omnibus estimates may conceal greater differences between specific groups of subjects. To address this question, we compute predicted SDR scores for each group in each List, using estimates from Tables 3 to 5. For this purpose, we

[29] In this context, no multivariate analysis is required to get the ACT, LISTT and LISTB estimates: they are simply the overall proportions (or mean numbers) of accepted items in the various treatments. However they may also be conceived of as means or proportions estimated from models in Tables 3 to 5, provided that all variables are kept at their mean values. We do not assume homogeneity of variances across treatments (an assumption that is rarely met in our data), and thus we use a corrected version of the t-test for independent samples, with degrees of freedom based on Satterthwaite approximation.

calculate the difference between the predicted LISTT and ACT scores; for example, predictions for male subjects in List 2 are 0.55 and 0.86, respectively, from which we obtain an SDR score of 0.31 (which, in this case, is not statistically significant).[30] Table 7 presents all SDR scores computed with this method. Clearly enough, cases of SDR are exceptions rather than the rule. More often than not, within-group comparisons of overt and covert responding do not provide sufficient evidence of SDR, at least from a statistical point of view. However, the few significant results in Table 7 are, as a whole, consistent with our hypotheses and lead to fairly plausible conclusions. We will focus here on findings from Lists 2 and 3 — we return to List 1 later on for comparison purposes.

To begin with, according to our **"non-involvement hypothesis"**, individuals with a more diffuse political profile should be more susceptible to SDR. In line with this expectation, *low levels of general political knowledge* are conducive to SDR in both Lists. When judging support for sensitive items from *direct* questions, the differences between poorly and highly knowledgeable subjects do not exceed 9 percentage points across Lists, but the difference is markedly higher in the *list* estimates (ranging between 40 and 84 percentage points). This suggests that the very high level of support for the job nondiscrimination and tax fairness statements which was demonstrated in answers to direct questions can be seen as a true endorsement of these principles on the part of knowledgeable subjects, but *not* on the part of less knowledgeable subjects. The latter probably lack the self-confidence to speak out their egoistic or anti-foreigner gut feelings. Interestingly, it is perhaps not by chance that *general* political knowledge appears to moderate the magnitude of SDR to some extent, whereas *specific* knowledge about welfare state issues and daily news consumption do not (with the exception of specific knowledge in List 2). Better than other measures, general political knowledge may reflect an overall subjective feeling of competence and legitimacy to express personal opinions. This disposition, which broadly corresponds to Bourdieu's (1979) and Gaxie's (2007) "statutory competence", was hypothesized to inhibit self-presentation biases.

The same line of reasoning applies to the role of age in List 3, where younger subjects appear more susceptible to SDR than their elders. Again, younger subjects overtly approved of liberal positions at rates comparable to (or higher than) older subjects, but their list counts were estimated to be clearly lower. All in all, this suggests that younger subjects tended to hide their true feelings about the tested items when given the opportunity to do so. Similar expectations hold with respect to the personal salience of welfare state issues. When relatively indifferent toward such issues, most individuals may not want to meet the costs of social opprobrium and are induced to disguise their true feelings. This is what we found for the tax fairness issue (List 3): SDR was more common among subjects who do not care much about welfare state issues. But, interestingly, the opposite relationship seems to be true when involvement was arguably least relevant, i.e., on the job nondiscrimination issue (List 2).[31] Finally, sex does not seem to play much of a role in SDR, at least with respect to welfare issues.

---

[30] The standard error of the difference between the LISTT and ACT predicted scores is computed simply as the square root of the sum of score variances. As we cannot assume homogeneity of variances across treatments, we use a corrected version of the t-test for independent samples, with degrees of freedom based on Satterthwaite approximation.

[31] Let us assume that, in List 2, the huge difference in SDR estimates between low- and high-salience subjects contains a grain of truth. It might then be argued that, unlike in List 3, welfare state matters were not *directly* relevant to judging the job nondiscrimination issue, but that caring about the welfare state generated *ambivalence* by promoting egalitarian values and setting up a clash with self-interest. In this sense, issue salience may have played a different role in List 2 than it did in List 3, and this role may have more to do with our *ambivalence* hypothesis.

Table 7: Agreement with the test items and SDR estimates in various subgroups, Lists 1–3 (187 ≤ N ≤ 191)

	LIST 1 (WELFARE ENTITLEMENT)							LIST 2 (JOB NONDISCRIMINATION)							LIST 3 (TAX FAIRNESS)							
	X1: List Experiment		X2: Direct Question		SDR estimate (X2 – X1)			X1: List Experiment		X2: Direct Question		SDR estimate (X2 – X1)			X1: List Experiment		X2: Direct Question		SDR estimate (X2 – X1)			
Subjects' characteristics	Mean	S.E.	Mean	S.E.	**Mean diff.**	S.E.	p (2-tailed)	Mean	S.E.	Mean	S.E.	**Mean diff.**	S.E.	p (2-tailed)	Mean	S.E.	Mean	S.E.	**Mean diff.**	S.E.	p (2-tailed)	Significant contrasts (p≤.10, 2-tailed)
Low general knowledge	0.625	.148	0.942	.047	**0.317**	.155	.043	0.326	.241	0.895	.063	**0.569**	.249	.024	0.542	.186	0.915	.061	**0.373**	.196	.058	LIST 1: low vs. high (p<.05)
High general knowledge	1.088	.137	0.979	.028	**-0.109**	.140	.435	1.165	.258	0.806	.098	**-0.359**	.276	.194	0.940	.192	0.852	.092	**-0.089**	.212	.677	LIST 2: low vs. high (p<.02)
Low specific knowledge[c]	0.795	.148	0.922	.058	**0.126**	.159	.428	0.369	.225	0.819	.076	**0.450**	.237	.059	0.683	.185	0.845	.077	**0.163**	.200	.418	none
High specific knowledge	0.843	.117	0.986	.020	**0.143**	.119	.230	0.970	.241	0.912	.064	**-0.058**	.250	.816	0.735	.176	0.939	.059	**0.204**	.185	.272	
Daily information < 30 min.	0.743	.141	0.971	.033	**0.228**	.145	.118	0.360	.231	0.728	.083	**0.369**	.246	.135	0.760	.196	0.894[a]	.052	**0.134**	.202	.509	none
Daily information ≥ 30 min.	0.910	.168	0.940	.052	**0.030**	.176	.864	1.053	.325	0.959	.037	**-0.094**	.328	.774	0.651	.206	0.894[a]	.052	**0.243**	.212	.255	
Low salience of WS issues	0.855	.149	0.941	.047	**0.086**	.156	.581	1.261	.250	0.862	.069	**-0.400**	.259	.125	0.432	.200	0.867	.078	**0.435**	.215	.044	LIST 2: low vs. high (p<.005)
High salience of WS issues	0.785	.140	0.974	.029	**0.189**	.143	.188	0.097	.299	0.867	.068	**0.771**	.306	.013	0.988	.160	0.915	.055	**-0.073**	.169	.667	LIST 3: low vs. high (p<.07)
Up to 21 years	0.487	.191	0.972	.035	**0.485**	.194	.014	0.688	.290	0.716	.120	**0.028**	.314	.928	0.466	.208	0.940	.043	**0.474**	.213	.027	LIST 1: 21– vs. 24+ (p<.01)
22-23 years	0.914	.109	0.988	.019	**0.074**	.111	.506	0.441	.284	0.867	.081	**0.426**	.296	.151	0.738	.224	0.696	.137	**-0.042**	.262	.873	LIST 3: 21– vs. 24+ (p<.03)
24 years and more	1.125	.170	0.852	.089	**-0.273**	.192	.160	0.944	.267	0.940	.051	**-0.003**	.271	.990	0.980	.225	0.696	.137	**-0.284**	.263	.283	
Man	0.811	.130	0.961[a]	.032	**0.150**	.133	.263	0.550	.235	0.865[a]	.054	**0.314**	.241	.194	0.676	.173	0.924	.050	**0.248**	.180	.171	none
Woman	0.835	.162	0.961[a]	.032	**0.126**	.165	.447	0.903	.267	0.865[a]	.054	**-0.039**	.272	.887	0.768	.218	0.824	.099	**0.056**	.240	.815	
Extreme left	0.847	.174	0.926	.080	**0.080**	.192	.681	0.779	.323	0.831	.068	**0.052**	.330	.876	0.525	.241	0.990	.019	**0.465**	.241	.056	LIST 3: extreme left vs. moderate left (p<.05)
Moderate left	0.947	.131	0.979	.024	**0.032**	.133	.809	0.454	.272	0.831	.068	**0.376**	.280	.181	1.019	.199	0.838	.078	**-0.181**	.214	.398	
Center/right	0.558	.301	0.935	.077	**0.377**	.310	.227	0.933	.426	0.931	.062	**-0.002**	.430	.996	0.428	.341	0.678	.213	**0.250**	.402	.534	
Discipline: political science	0.600	.178	0.984	.020	**0.384**	.179	.034	0.636	.227	0.893	.056	**0.257**	.234	.274	0.417	.231	0.886	.090	**0.469**	.248	.060	LIST 2: economics vs. other (p<.09)
Discipline: economics	1.042	.217	0.984	.020	**-0.059**	.218	.788	0.248	.459	0.893	.056	**0.645**	.462	.165	1.121	.306	0.859	.135	**-0.262**	.334	.434	
Discipline: history	0.773	.326	0.984	.020	**0.210**	.327	.521	0.148	.734	0.893	.056	**0.745**	.736	.314	0.427	.397	0.768	.165	**0.340**	.430	.430	LIST 3: political science vs. economics (p<.08)
Discipline: other	0.874	.205	0.855	.088	**-0.020**	.223	.930	1.112	.292	0.807	.094	**-0.305**	.307	.322	0.793	.242	0.936	.057	**0.143**	.249	.566	
Swiss citizen	0.779	.115	0.961[a]	.032	**0.182**	.119	.129	0.567	.197	0.865[a]	.054	**0.298**	.205	.147	0.877	.146	0.894[a]	.052	**0.017**	.155	.913	LIST 3: Swiss vs. foreigner (p<.03)
Foreigner or binational	0.962	.174	0.961[a]	.032	**0.000**	.177	.998	1.073	.315	0.865[a]	.054	**-0.208**	.319	.516	0.125	.304	0.894[a]	.052	**0.769**	.309	.014	
Belief activation: no	0.879	.123	0.933	.052	**0.054**	.134	.687	0.424	.241	0.883	.062	**0.459**	.249	.067	0.631	.181	0.868	.072	**0.237**	.195	.226	LIST 2: no vs. yes (p<.09)
Belief activation: yes	0.751	.117	0.978	.025	**0.227**	.120	.060	0.978	.228	0.843	.075	**-0.135**	.240	.575	0.801	.196	0.915	.057	**0.114**	.204	.578	

Notes: Predicted values are based on models in Tables 3–5. [a]: Groups for which no Direct Question estimate is available are attributed the estimated means from the PML models (i.e., the predicted ACT value when all variables are set to their means); these estimates differ from the aggregate proportions displayed in Table 6 because, compared to the difference-in-means approach, the multivariate approach is based on a smaller sample due to missing values in some predictors.

Overall, we find some supportive evidence that SDR is more widespread among individuals with less cognitive political resources, and almost no evidence pointing to the opposite relationship. The part played by "cultural capital" in promoting tolerance and equality, as shown in many studies, does not seem to be undermined by social desirability concerns (for similar conclusions, see Heerwig and McCabe 2009).

Turning now to our **"ambivalence hypothesis"**, SDR is expected to prevail among people who hold ambivalent attitudes about an issue. In the context of this study, we assume that ambivalence stems largely from a conflict between values and affective orientations, on the one hand, and utilitarian, selfish goals, on the other hand. Which ingredients of ambivalent attitudes will dictate overt responses to some sensitive issue is expected to depend on the mix of social identifications and ensuing social norms which are made salient by the context and by the specific items. In this regard, three types of identification groups may be primed by the questions: ideological groups, membership groups (in this context, the student community and the lower-level curriculum groups), and national community. These were assessed by left-right self-placement, academic discipline, and the distinction between Swiss citizens and foreigners/multiple citizens, respectively. Finally, the procedure of "belief activation" is assumed to foster ambivalence, because subjects are required to list both the pros and cons of the welfare state issue as they perceive it.

The particularity of most of these variables is twofold. First, they are associated both with the likelihood of genuinely disagreeing (or agreeing) with some issue *and* with the likelihood of providing a socially desirable response, given disagreement; this allows for the possibility of nonlinear relationships between moderators and SDR. Second, the expected effects of these variables strongly depend on the response context, and especially on the nature of items. Left-right orientations can serve to illustrate these two aspects. In this study, all three tested items might be characterized as social-democratic or leftist, or even "socialist". Subjects who define themselves politically as "far left" or "moderate left" likely share the values underlying these items. Endorsement of these values, in turn, is reinforced by identification with ideological groups ("the left", "Social Democrats", or "anti-capitalists") and/or with the "campus community", which is usually tilted toward left values.[32] *When*, however, left-wing subjects *do* disagree with the items, in particular because self-interest prevails over affect-laden moral considerations, they may feel greater pressure to conceal their "true" attitudes than do other student groups. This is because peer group influence combines with self-esteem goals to provide strong social desirability incentives. Conversely, right-wing subjects are more likely to disagree with the proposed items. At the same time, they may feel less pressure to conform to the majority view because their values and self-interest coincide. In addition, right-wing subjects may enjoy support for their ideas from other reference groups (including primary groups such as family and peer groups from the subjects' own faculty or curriculum) which insulates them from the larger influence of the student community or misleads them into believing that their opinion is prevalent among students. All in all, which ideological group will show the greatest propensity for social desirability is an empirical question.[33]

---

[32] Indeed there was a strong left bias in our students' sample. On a 0-10 left-right scale, the average self-placement in the students sample (excluding apprentices) was 3.3 (median=3), and more than 80 percent were posited to the left of the 0-10 scale midpoint. However, it may again be noted that left-right self-placement is likely to be itself affected by SDR.

[33] To some extent, similar considerations apply to a comparison of subjects' academic disciplines and nationalities. As contrasted to economics and history, students in political science may be more susceptible to SDR because of their presumed higher endorsement of universalistic values, political correctness, and rejection of utilitarian motives of political action. Foreigners (including "binationals", i.e., Swiss citizens holding a second nationality) may feel particularly committed to the policy goals embodied in our test items (at least in Lists 1 and 2), since these policies may be

As shown in Table 7, subjects' self-reported ideology mattered only with respect to the tax issue (List 3). Extreme left students were unanimous in supporting tax fairness when asked directly about it; however, taking covert responses into account, they appear significantly more susceptible to SDR than moderate left students. Nothing comparable occurs in List 2, where all three ideological groups have similar SDR scores. Thus, there is only partial evidence to support our hypothesis about the role of left-right orientations in SDR. Likewise, the role played by our subjects' academic discipline seems rather modest. As hypothesized, political scientists *do* appear particularly susceptible to SDR with respect to the tax fairness issue (an estimated 47 percent of them hide their true feelings about this issue). But none of the other groups shows a significant propensity to SDR.[34] Importantly, our multivariate framework allows us to rule out the possibility that political scientists display a heightened susceptibility to SDR simply because they make up a disproportionate share of extreme left-wingers.[35]

The nationality of our subjects (i.e., the distinction between Swiss and foreign/ multiple citizens) seems to matter in some circumstances. However, SDR was detected where we least expected it, namely for the issue of tax fairness (List 3) — and *not* on the job nondiscrimination item. On the tax issue, the SDR estimate for Swiss subjects is virtually 0, while it is nearly 80 percent and highly significant ($p<.02$) for foreigners and multiple citizens! Only about one tenth of foreigners actually supported the tax fairness policy goal in covert conditions, as compared to 90 percent supporting it in their answers to the direct question. Arguably, opposing fiscal redistributive policies can be perceived as a threat to the welfare of minorities which are the main recipients of these policies. As a matter of fact, foreign students have better life chances than most members of their national community. Therefore, rejecting wealth redistribution on ideological or utilitarian grounds and failing to live up to moral obligations toward one's community may arouse feelings of guilt and betrayal leading to SDR.[36]

Finally, turning to the role of belief activation in SDR, our analysis yields inconsistent results. On the one hand, these results suggest that subjects who answered the likes/dislikes questions were either no different from those who did not (List 3) or *less* susceptible to SDR (List 2). In fact, answers to the job nondiscrimination item reveal a substantial contrast between the "activation" and "no activation" groups which is exactly the *opposite* of that implied by our ambivalence hypothesis. That is, subjects who were asked to express their pros and cons toward the welfare state were clearly *less* susceptible to SDR than subjects who were not instructed to do so.

---

seen to profit first and foremost the foreign labor force and other socially disadvantaged groups. However, when ambivalent about the test items, notably because their own interests are perceived to conflict with their salient values and allegiances, foreign students may downplay utilitarian considerations in their overt answers, but probably not in their list counts.

[34] On the tax fairness issue (List 3), political science students are significantly different from economics students, as the latter overwhelmingly supported the welfare entitlement item in both overt and covert answering conditions. The other significant contrast (between economics and "other" disciplines in List 2) distinguishes groups whose SDR scores, by themselves, do not differ from zero at conventional levels of statistical significance. History students have SDR scores comparable to their political science counterparts; however, the small size of this group drives these tendencies away from statistical significance.

[35] In fact, there are more left-wingers in disciplines where SDR estimates are lower than for political scientists (e.g., history, geography, literature and linguistics, and other human sciences).

[36] Alternatively, it may be argued that the *principle* of tax fairness is supported by foreigners in their direct responses, but that the *implementation* of the principle is perceived as unfair to one's family (assuming its revenues are modest) and thus rejected in the item counts. This kind of "policy-implementation" gap is a staple of public opinion analysis which is amenable to our ambivalence hypothesis, since principle and implementation have different implications for public values and private interests.

One possible explanation for this finding is that the likes/dislikes questions about the welfare state are not well tailored to opinions on the job nondiscrimination issue and that attitudinal ambivalence may be inconsequential for SDR on this particular issue. Instead, asking subjects about their likes and dislikes, that is, about matters of *personal* taste and judgment, may bring about a *self-individuating process*, as shown by the use of similar experimental procedures (e.g., Ambady et al., 2004). Consistent with the notion of an individuation–categorization continuum (Fiske et al., 1999), inducing subjects to focus on their own beliefs and motives could *deemphasize group identities*. Accordingly, self-focused subjects might attach less importance to social expectations. However, if this reasoning has some validity, it remains unclear why it does not extend to the other tested issues.

To what extent does evidence obtained for List 1 ("welfare entitlement" item) provide confirmation of the above findings? Remember that, due to the presence of "ceiling liars" (see section 5.1), results for this list are to be taken with extreme caution; however they may be used as auxiliary evidence about the role of moderators insofar as they reproduce findings obtained in other lists. By and large, results from List 1 square well with the evidence discussed above. To begin with, these results are consistent with the hypothesis (supported in the two other lists) that SDR is associated with low levels of general knowledge. Likewise, younger subjects and political science students may be especially prone to SDR, which is consistent with results from List 3. Further, political ideology seems unrelated to SDR in List 1 — like in List 2, but unlike List 3; this would tend to buttress the idea that ideology is at stake in providing socially desirable answers to some kinds of issue, but not to others. Finally, our analysis of List 1 suggests that subjects whose beliefs were "activated" were indeed more susceptible to SDR. Although this is in line with our ambivalence hypothesis, we may recall that the results from List 2 suggested an opposite relationship between belief activation and SDR. Accordingly, it seems safe to say that no solid evidence can be adduced in favor of the belief activation hypothesis.

5.3. Comparison with the "difference-in-means" method

We have analyzed the present data using the more classical "difference-in-means" (i.e., bivariate) approach briefly exposed in section 4.2. This allows us to compare group estimates found using either method — bivariate or multivariate — and to determine whether SDR values based on the difference-in-means approach are at least good approximations of SDR values obtained with more sophisticated methods. Because the ceiling effect in List 1 examined above will appear regardless of the analytical method, we keep all three lists in this comparison.

Averaging over all groups and Lists (i.e., 24 groups × 3 lists), overall SDR scores look much the same for the bivariate vs. multivariate method (0.13 vs. 0.16). However, these average values conceal much of the variation observable within particular groups. Thus, the typical difference in a group amounts to 0.14, and it is about twice as large with respect to the job nondiscrimination item (List 2) as it is for the two other items (0.21 vs. 0.10). In addition, bivariate estimates typically have less sampling variability and are more precise.[37] Accordingly, smaller SDR estimates from the bivariate method may emerge as statistically significant, whereas larger SDR estimates from the multivariate method may not.

All told, the differences in empirical results and conclusions drawn from either empirical approach are not overwhelming. All but five effects reported in Table 7 were also obtained with the bivariate approach ($p<.10$). The other way round, nine

[37] Averaging over all groups and Lists, the standard error of our estimates amounts to 0.19 in the bivariate framework, as against 0.24 in the present study (average p-values are 0.22 and 0.40, respectively).

of the 12 effects uncovered by the bivariate method also emerge in the multivariate analysis.[38] In sum, there is admittedly a good deal of overlap between empirical results from bivariate and multivariate analyses. But if one considers the multivariate strategy followed in this paper a more reliable method to detect SDR tendencies, then our comparison suggests that the use of the bivariate method increases the risk of producing both "false negatives" (overlooking actually existing relationships) and "false positives" (uncovering individual characteristics that purportedly enhance SDR when in reality they do not).

Therefore, the multivariate approach recommended in recent years is certainly the better choice. For another thing, it allows researchers to take into account omitted variable bias and to rule out a number of spurious relationships. Take for example the relative SDR estimates of male and female respondents in List 2. Bivariate analysis suggests that 45 percent of women and 0 percent of men (both estimates with large standard errors) made socially desirable responses to the job nondiscrimination item. This significant contrast ($p<.08$), however, does not pass a multivariate test (see Table 7). If anything, it appears that *men* were possibly susceptible to SDR ($p<.20$) while women were not — even though this gender difference is not significant. What probably accounts for this reversal of findings is the lack of independence between factors explaining SDR tendencies. As it turns out, gender is strongly related to general political knowledge ($\chi^2=22.3$, $p<.001$; $\Phi=.33$), which is one of the most important predictors of SDR on this item. It may thus be argued that male-female differences in SDR are mostly attributable to this covariate, and not to gender per se. Accordingly, the impact of gender disappears when SDR is studied in a multivariate framework.

5.4. Little lies and some bad conscience?

Unlike most list experiments whose results have been published so far, the CTRL group in our experiment was asked direct questions on *all* items, i.e., on test *and* baseline items. This design feature enables us to compare the number of baseline items accepted via the direct questions in the control group with the number of baseline items accepted via the short list count.[39] The difference between these two values provides an estimate of the extent of SDR on *baseline* items. (Again, we draw attention to the fact that covert answers to baseline items in List 1 show evidence of a ceiling effect; accordingly, the difference between item counts and "pseudo" item counts from overt questions is artificially reduced, and findings from this list should be interpreted with caution). Building on this, we ask two questions that may shed additional light on the nature and mechanisms of SDR. First, is there SDR on baseline items too? Second, if there is evidence of SDR on baseline items, is it related in any way to SDR on test items?

The answer to the first question is a qualified "yes". Collapsing all subgroups and focusing on SDR on *baseline* items, the overall difference between "overt" and "covert" conditions is negligible in Lists 1 and 3 ($\Delta = -0.04$ and 0.05, respectively; both $ps>.40$), but in List 2 it is substantial and highly significant ($\Delta = 0.30$; $p<.01$). This suggests that we should not assume a priori that SDR only occurs with respect to "sensitive" items (see also Zigerell, 2011). Put another way, it is certainly worth

[38] By "effects" we mean here all significant within-group differences between direct answers and list counts, either in the positive or (much more rarely) in the negative range.

[39] Estimates were obtained in exactly the same way as for the acceptance of test items. Multivariate analysis using the ICTREG procedure was applied for the estimation of baseline counts from the ITEM3 and ITEM4 groups; OLS multiple regression was used to predict the number of accepted items in the CTRL group.

asking whether baseline items in our Lists are really as "nonsensitive" as they may seem, and which factors contribute to untruthful answers on these items.[40]

The second question invites us to compare SDR by the same groups in reaction to different categories of items. From a theoretical perspective, three types of relationships between SDR on test items and SDR on baseline items can be expected. First, there may be no relationship at all. For example, those who are prone to SDR on one type of items are just as likely to show a positive, negative or nonexistent SDR score on the other type of items — in a nutshell, SDR on test items and SDR on baseline items are entirely distinct phenomena. Second, the two SDR measures may be positively related. This is what should be expected if, at least for a substantial proportion of subjects, SDR originates from a *stable disposition* (e.g., a personality trait) rather than from contextual cues related to particular circumstances. Third, the two measures may be negatively related. In this case, compensatory mechanisms may be at work, such that individuals who "lie" on one item (either test or baseline) *consciously or unconsciously avoid lying* on other items.

Figure 1 gives an overview of SDR on both test and baseline items in all three Lists and for each subgroup of subjects. In practical terms, all groups for which confidence intervals are in the positive range and do not include the zero value are likely to comprise a significant proportion of "liars" ($p<.10$). Figure 1 allows us to draw a number of interesting conclusions. First of all, many respondents "lied" on baseline items, too! Across the various Lists (but in particular in List 2), no less than thirteen categories of subjects display a positive and statistically significant SDR score on baseline items, and eight categories have an estimated average SDR score of 0.5 or higher. For example, in List 2, subgroups such as politically knowledgeable, older, extreme left, or foreign/naturalized subjects were estimated to overreport their agreement with baseline statements by about 0.5 items or more (all $ps<.05$). More importantly, another striking feature of Figure 1 is that SDR on test items and SDR on baseline items do not appear to follow the same pattern. More often than not, categories of subjects showing a high propensity to engage in SDR on the test item do not show a similar tendency with respect to baseline items. The other way round, most categories appearing to conceal their true feelings on baseline items are found to be "sincere" on test items.[41] Overall, then, these results are suggestive of a negative relationship between SDR on test items and SDR on baseline items.

Because we are dealing with aggregate results, we should beware of fallacious inferences.[42] Therefore, we conducted another analysis that goes some way toward *individualizing* SDR estimates, enabling us to test whether the interplay of social desirability mechanisms on test and baseline items can be explained at the individual level. This analysis is premised on the fact that subjects in our experiment are characterized by multiple attributes (such as age or ideology), each of which can be associated with heightened or reduced social desirability propensities (see Figure 1). Accordingly, we use the models presented in Tables 3 to 5 to predict the net result of all individual characteristics on each subject's SDR score. Regardless of which

[40] At this point, it is important to recall that the existence of SDR on baseline items does *not* imply a violation of the "no design assumption" examined above. However, in the event where SDR on test items and SDR on baseline items are systematically related, the question of whether respondents give dishonest answers on baseline items deserves further examination.

[41] Overall, there is just *one* category that engaged in SDR on *both* types of items ($p<.10$), namely subjects who received "no belief activation" in List 2.

[42] In particular, we cannot assume that the same individuals were the driving force of aggregate tendencies, and that those who gave untruthful answers on one type of items (say, test items) were mostly honest on the other type (say, baseline items) — although it may sometimes be the case.

treatment group subjects were assigned to, their characteristics are used to yield predicted values from which we derive *individual* social desirability scores on the test item and on baseline items: $\sim\mathrm{SDRt}_i = \widehat{\mathrm{ACT}}_i - \widehat{\mathrm{LISTT}}_i$, and $\sim\mathrm{SDRb}_i = \widehat{\mathrm{ACB}}_i - \widehat{\mathrm{LISTB}}_i$. We compute predictions for the whole set of subjects for whom we have valid observations on all variables, since they reflect the full distribution of variables from which the regression coefficients were generated. Properly speaking, SDR scores are not merely "estimated" in this way, but rather "*simulated*" in the sense that each individual score (say, for a subject in the CTRL group) is based, in part, on individual responses obtained in *other* experimental conditions (say, from subjects in the ITEM3 and ITEM4 treatment groups). Accordingly, the tilde symbol in ~SDRt$_i$ and ~SDRb$_i$ scores stands for the fact that these individual scores are simulated. In this sense, our method breaks from conventional inferential statistics and must be considered a heuristic tool for broadly comparing SDR measures which are otherwise unavailable on a purely individual basis. Again, because the "individual" scores are in fact a compound of information gathered from many subjects, we cannot rule out that the relationship between ~SDRt$_i$ and ~SDRb$_i$, interpreted here heuristically as if scores were measured within single individuals, is a mere illusion of the simulation framework. With this caution in mind, we show in Figure 2 the relationship between simulated ~SDR scores on both types of items.

Figure 2 largely confirms what the analysis of subgroups at the aggregate level has suggested. There is a moderate, negative, and not perfectly linear relationship between ~SDRt$_i$ and ~SDRb$_i$. It is also substantially stronger for List 2 than for Lists 1 and 3; in general, though, the same relationship holds for all Lists. To be sure, there is a good deal of uncertainty in the data displayed in Figure 2, owing to cumulative prediction errors at each estimation step and to the simulation framework underlying the figure. Still, the overall pattern in Figure 2 is consistent with the aggregate analysis discussed earlier and cannot be dismissed so easily. When SDR on test items is high, the odds are that SDR on baseline items is relatively limited. Conversely, when SDR on test items is low, SDR on baseline items tends to be more pronounced.

Building on these results, we now consider what individual-level mechanisms might account for the kind of "sincerity-deceit" trade-off suggested in Figures 1 and 2. As a first working hypothesis, we may assume that individuals do not have unlimited tolerance for their own deceptive behavior. Confronted with a list of four items, and without any suspicion that one of them is the target of the experiment, subjects may be tempted to hide their true feelings on some of these items. Beyond some threshold, however, the psychological cost incurred from perceiving oneself as a liar may exceed the satisfaction experienced from complying with social norms. For many individuals, so the hypothesis goes, the acceptable limit amounts to a "single little lie". Accordingly, subjects who provide an untruthful answer to the test item should feel compelled to give sincere answers to all other (baseline) items; the other way round, subjects who are honest on the test item should find it tolerable to conceal their true opinion about one of the baseline items. If at least a substantial minority of subjects conforms to this pattern, it follows that ~SDRt$_i$ and ~SDRb$_i$ should be inversely related.

Figure 1: Estimates of SDR (●: test items; ●: baseline items) with 90% confidence intervals

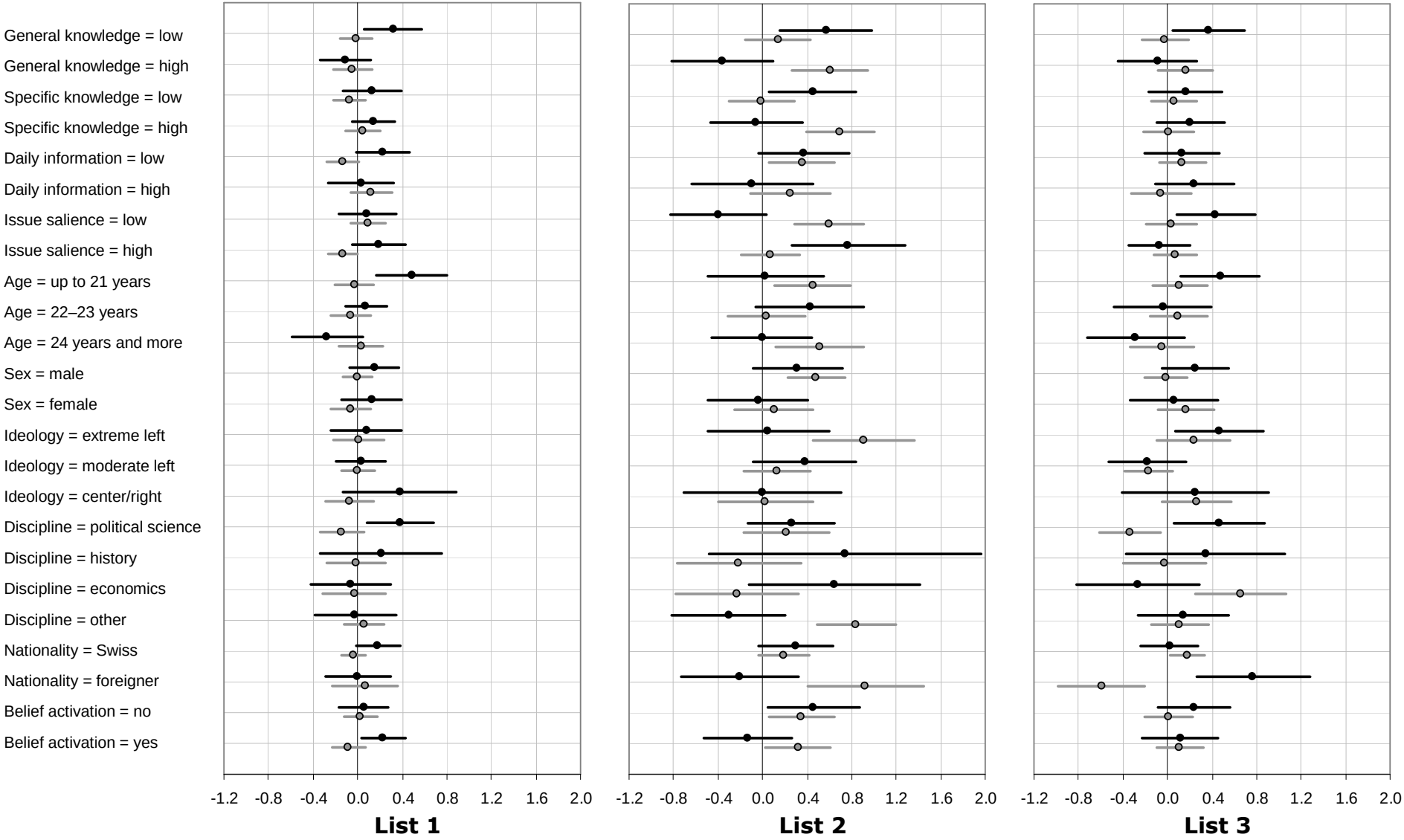
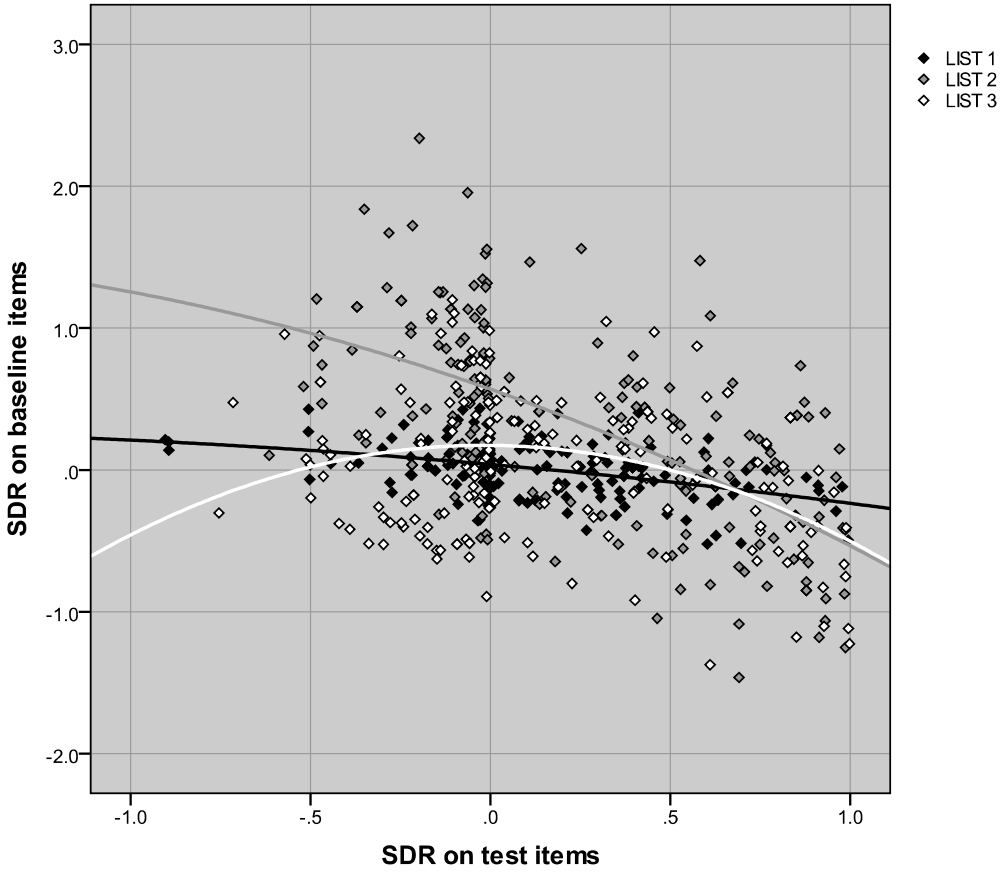


Figure 2: The relationship between SDR on test items and SDR on baseline items



List 1: $\hat{y} = -0.0515x^2 - 0.2229x + 0.0394$ ($R^2 = 0.187$)
List 2: $\hat{y} = -0.2089x^2 - 0.8945x + 0.5690$ ($R^2 = 0.368$)
List 3: $\hat{y} = -0.6532x^2 - 0.0220x + 0.1736$ ($R^2 = 0.138$)
Note: predicted values for LISTT were truncated for values higher than 1 and lower than 0.

This introspective mechanism may explain why social desirability does not weigh equally on responses to item and baseline items. However, it does not account for the variations between groups depicted in Figure 1. In this respect, two main factors may be distinguished. First, some characteristics may be inversely related to $\sim SDRt_i$ and $\sim SDRb_i$ for largely incidental reasons having to do with our particular choice of baseline items.[43] Second, and related to the previous point, groups probably vary in their *perceptions of the sensitiveness of baseline items*. For example, politically aware individuals may see a policy issue as complex, multidimensional and potentially contentious, and hence may be reluctant to take sides, while other individuals may not see much of a problem in the issue. This may be especially the case with the more "technical" issues raised in List 2. Alternatively, some categories of subjects may be disproportionately undecided *and* unaware of the social norm on some of the baseline issues. Subjects from such categories may tend to choose more often the DK option when responding to direct questions, but we would expect them to follow some rule of thumb when answering the list questions. This is a question that has not received attention from researchers using list experiments (Lax et al., 2014: 19), and it is unclear whether items on which subjects are unable to take position tend to be counted as "accepted" or "refused". Although it may contribute to explain the pattern of results shown in Figure 2, the issue of indecision and DK answers requires a deeper examination. To investigate this question, we provided subjects in the CTRL group with the possibility to pick a "don't know" answer (DKA) to the direct questions. We reasoned that such answers may originate either in strategic reasons (to disguise a socially undesirable opinion) or in true indecision or ambivalence. 21% of subjects provided *at least one* DKA to baseline items in List 2, as against 3% in List 1 and 1% in List 3. Our analysis of List 2 (not shown here, but available upon request) suggests that DKAs hint at true indecision, rather than being *fake* answers designed to conceal socially reprehensible opinions. It also indicates that indecision is related to a heightened propensity to count obscure items as "accepted" in list conditions.

6. Conclusion

In this paper, we have assumed that social desirability responding (SDR) occurs about as frequently in matters of welfare policy as it does in many other fields. In particular, the statement that "taxes should be raised more equitably, even if I or my family were to be taxed more heavily" appears to have elicited biased responses from a non-negligible proportion of subjects. These might have felt that refusing to contribute directly to more equality in the distribution of the tax burden would seem too selfish or mean-spirited. On the other hand, the proposal was sufficiently unattractive to warrant refusal when it could go unnoticed.

Recent advances in the "List Experiment" method allowed us to conduct a multivariate analysis of SDR moderators. In this analysis, three variables stand out as particularly relevant for explaining variations in SDR scores: political knowledge, age, and academic discipline. To begin with, subjects with poor political knowledge were especially likely to overreport their approval of the welfare entitlement, tax

[43] For example, in List 2, politically aware subjects appear truthful in their support for equality and social justice (as encapsulated in their answers to the test item), but they may be more reluctant to voice their opposition to the "popular sovereignty" argument raised in baseline items 1 and 2 on matters of European integration (see Appendix A). The other way round, subjects with less political knowledge were shown to be rather dishonest on test items (in all three Lists). At the same time they may be more prone to adhere to the "simplistic" view that the popular will should dictate policymaking and that all matters of foreign policy are to be settled by a popular vote; hence, they may approve baseline items without much afterthought.

fairness, and job nondiscrimination items; younger subjects were also inclined to do so, except for the last item. We interpret these findings as evidence that individuals with *less involvement toward politics* are less committed to their own opinions and less willing to "publicly" maintain a socially undesirable position. Uninvolved individuals are thus more likely to go where the wind blows — in this case the wind of the "campus community". In contrast, the non-involvement argument falls short of explaining why political science students were *more* susceptible to SDR (at least on two of the three issues) than were subjects from other disciplines. Instead, our *ambivalence hypothesis* posits that individuals who have conflicting considerations about an issue are more likely to comply with social norms. We assume that the self-interest of many political science students collided with their deeply entrenched values or beliefs toward social matters, which increased the gap between "covert" and "overt" opinions. The fact that academic discipline was more consistently associated with SDR than other variables supposed to foster ambivalence (ideology, nationality, belief activation) is not entirely surprising. To some extent, political science students should be better aware of the *majority* (i.e., *"politically correct") opinion* than students from other curricula, not least because they are expected to talk frequently about politics with other people, both within and outside the campus community.

On a more technical note, we have shown that list experiments can be successfully applied to ordinary issues reaching beyond the usual "gut responses" and "anger items" studied in early research using this method. Simple paper-and-pencil questionnaires are fit for purpose. Likewise, our research suggests that various types of baseline items may be used (including those which themselves elicit SDR), as well as test items which are almost universally accepted. However, despite its promises, the present research also has a number of important limitations. First, unlike the survey experiments used in most list experiment research, our study combines the disadvantages of small-N quantitative analysis (unstable and inaccurate estimates, under-detection of effects, etc.) with those of controlled experiments — most particularly *low external validity*. Our results do not readily generalize to other student communities, let alone to the population of Switzerland or other countries. Second, our tests of the ambivalence and non-involvement hypotheses are admittedly very crude. Moderator variables were subjectively chosen because they appeared to be good proxies for these two key factors (e.g., political knowledge for measuring the non-involvement dimension), but also because they allow for a comparison with other studies using the same (rather standard) variables. In any case, it should be the task of future investigations to develop direct and internally valid measures of ambivalence and non-involvement.

Appendix A: Wording of items in all lists

List 1:

- Employment offices should help the unemployed to find a new job. (% agreeing in CTRL group: 93.0)
- **Foreigners should have the same rights to social services as Swiss citizens, if they have lived and worked a long time in Switzerland. (94.3)**
- Women should not be entitled to social benefits, they can always become housewives. (0)
- The state should ensure a decent living standard to the needy elderly. (94.3)

List 2:

- Whenever the issue arises, the Swiss people should have the right to decide whether they want Switzerland to join the EU. (94.1)
- The bilateral agreements concluded with the EU do not give the EU the right to dictate Swiss policy toward third countries (i.e., non-EU members). (84.7)
- As a compensation for joining the EU, we may accept to give up important elements of our advanced transport policy (e.g., heavy vehicle fee, protection of the Alps, transfer of goods transport from road to rail). (16.2)
- **Foreign job seekers with equal skills and experience should not be discriminated against in favour of Swiss job seekers. (81.7)**

List 3:

- Someone who has not contributed to pension schemes (old age and survivors insurance, occupational pension funds) should not receive a pension or social benefits at retirement age. (30.0)
- The state should provide primary health care services. (93.0)
- **Taxes should be raised in a more equitable manner, even if I or my family were to be taxed more heavily. (88.1)**
- No full-time worker should live under the so-called poverty line. (98.6)

Appendix B: Construction of moderator variables

General political knowledge

General political knowledge was measured using a set of factual knowledge questions. First, four questions were asked to measure the subjects' knowledge of fundamental *institutions* of the Swiss political system: (a) popular initiative; (b) federalism; (c) the role of Parliament in electing the government members; and (d) the number of governing parties. One point was assigned for each correct answer (half points were also assigned to responses that were incorrect but not far off the mark). The scores were summed to produce a first knowledge scale (α = .71). Second, questions aimed to measure the subjects' knowledge of *political leaders*. A list of six leaders was provided, whose renown

was presumably either quite high (3 of them) or rather low (3 of them). The task was to associate each leader with his or her party among four possible options, i.e., the four main governing parties (SP, CVP, FDP, SVP). One point was assigned for each correct answer, and the scores were also summed to produce a second knowledge scale (α = .75). As the two scales are highly correlated (r = .65), they were standardized and summed to yield a single scale of general political knowledge, ranging from 0 to 10 (M=7.6, SD=2.4; 23% of subjects with highest score). The variable was dichotomized with a cut-off value of 8.5 (low=60 percent of subjects).

Issue-specific knowledge

Four questions were included in the questionnaire to gauge the subjects' knowledge of welfare policies (similar questions were asked about Swiss European policy, so as to prevent subjects from thinking that welfare issues were the focus of the experiment). For example, we asked which area of welfare state activity was introduced only in 2004 in Switzerland (the correct answer, maternity insurance, was to pick from a list containing three other response options: unemployment insurance, disability insurance, and compulsory health insurance.) One point was assigned for each correct answer; the scores were summed to produce a scale of issue-specific knowledge (α = .65), ranging from 0 to 4 (M=3.0, SD=1.2; 47% of subjects with highest score). The variable was dichotomized with a cut-off value of 8.0 (low=53 percent of subjects).

Daily information

News exposure was assessed by the following question: "On average, how much time a day do you spend following political news on national and international affairs?" The five response categories were combined into a low exposure (up to 30 min.; 58 percent of subjects) vs. high exposure (more than 30 min.) dichotomy.

Personal salience

The question reads: "How important to you personally is the issue of the role and performance of the Swiss welfare state?" An eleven-point scale was provided, from "very unimportant" (0) to "very important" (10), with a DK option. The variable was dichotomized with a cut-off value of 8.0 (low=51 percent of subjects).

Ideology

Ideology was tapped by a single question of left-right self-placement, ranging from 0 to 10, with a DK option. As one would have expected given the population from which our sample of subjects was drawn, the ideology measure is skewed toward leftist positions (M=3.6, SD=1.9), with some 74% of subjects positioning themselves to the left of the scale midpoint. The scale was partitioned into three categories: extreme left (positions 0–2; 30 percent of subjects), moderate left (3–4; 44 percent), and center/right (5–10; 26 percent).

References

Aarts, Henk, Ap Dijksterhuis and Ruud Custers (2003). "Automatic normative behavior in environments: the moderating role of conformity in activating situational norms", *Social Cognition* 21(6): 447-464.

Abrams, Dominic and Michael A. Hogg (1990). "Social identification, self-categorization and social influence", *European Review of Social Psychology* 1(1): 195-228.

Albarracín, Dolores and Amy L. Mitchell (2004). "The role of defensive confidence in preference for proattitudinal information: how believing that one is strong can sometimes be a defensive weakness", *Personality and Social Psychology Bulletin* 30(12): 1565-1584.

Albarracín, Julia, Wei Wang and Dolores Albarracín (2012). "Are Confident Partisans Disloyal? The Role of Defensive Confidence in Party Defection", *Journal of Applied Social Psychology* 42(7): 1576-1598.

Alwin, Duane F., Ronald L. Cohen and Theodore M. Newcomb (1991). *Political Attitudes over the Life Span: The Bennington Women after Fifty Years*. Madison: University of Wisconsin Press.

Ambady, Nalini, Sue K. Paik, Jennifer Steele, Ashli Owen-Smith and Jason P. Mitchell (2004). "Deflecting negative self-relevant stereotype activation: The effects of individuation", *Journal of Experimental Social Psychology* 40: 401-408.

Armingeon, Klaus, Fabio Bertozzi and Giuliano Bonoli (2004). "Swiss Worlds of Welfare", *West European Politics* 27(1): 20-44.

Arts, Wil and John Gelissen (2001). "Welfare States, Solidarity and Justice Principles: Does the Type Really Matter?", *Acta Sociologica* 44: 283-299.

Baldwin, Peter (1990). *The Politics of Social Solidarity: Class Bases of the European Welfare State 1875–1975*. Cambridge: Cambridge University Press.

Banting, Keith G. (2000) "Looking in three directions: migration and the European welfare state in comparative perspective", in M. Bommes and A. Geddes (eds.), *Immigration and Welfare: Challenging the Borders of the Welfare State*. London: Routledge (13-33).

Bassili, John N. (2003). "The Minority Slowness Effect: Subtle Inhibitions in the Expression of Views Not Shared by Others", *Journal of Personality and Social Psychology* 84(2): 261-276.

Bassili, John N. and Joseph F. Fletcher (1991). "Response-Time Measurement in Survey Research: A Method for CATI and a New Look at Nonattitudes", *Public Opinion Quarterly* 55: 331-346.

Bay, Anne-Helén and Axel West Pedersen (2006). "The Limits of Social Solidarity: Basic Income, Immigration and the Legitimacy of the Universal Welfare State", *Acta Sociologica* 49(4): 419-436.

Belli, Robert F., Michael W. Traugott and Matthew N. Beckmann (2001). "What Leads to Voting Overreports? Contrasts of Overreporters to Validated Voters and Admitted Nonvoters in the American National Election Studies", *Journal of Official Statistics* 17(4): 479-498.

Belli, Robert F., Sean E. Moore and John VanHoewyk (2006). "An experimental comparison of question forms used to reduce vote overreporting", *Electoral Studies* 25: 751-759.

Berinsky, Adam J. (1999). "The Two Faces of Public Opinion", *American Journal of Political Science* 43(4): 1209-1230.

Berinsky, Adam J. (2002). "Political Context and the Survey Response: The Dynamics of Racial Policy Opinion", *Journal of Politics* 64(2): 567-584.

Berinsky, Adam J. (2004). "Can We Talk? Self-Presentation and the Survey Response", *Political Psychology* 25(4): 643-659.

Bernstein, Robert, Anita Chadha and Robert Montjoy (2001). "Overreporting Voting. Why It Happens and Why It Matters", *Public Opinion Quarterly* 65(1): 22-44.

Blair, Graeme and Kosuke Imai (2012). "Statistical Analysis of List Experiments", *Political Analysis* 20(1): 47-77.

Blair, Graeme and Kosuke Imai (2014). *List: Statistical Methods for the Item Count Technique and List Experiment*. R package version 7.1.

Blair, Graeme, Kosuke Imai and Jason Lyall (2014). "Comparing and Combining List and Endorsement Experiments: Evidence from Afghanistan", *American Journal of Political Science* 58(4): 1043-1063.

Blekesaune, Morten and Jill Quadagno (2003). "Public Attitudes toward Welfare State Policies: A Comparative Analysis of 24 Nations", *European Sociological Review* 19(5): 415-427.

Bourdieu, Pierre (1979). *La Distinction. Critique Sociale du Jugement*. Paris : Minuit.

Bradburn, Norman N., Seymour Sudman and Brian Wansink (2004). *Asking Questions. The Definitive Guide to Questionnaire Design — For Market Research, Political Polls, and Social and Health Questionnaires*. New York: Wiley & Sons.

Brewer, Noel T., William K. Hallman, Nancy Fiedler and Howard M. Kipen (2004). "Why Do People Report Better by Phone Than by Mail?", *Medical Care* 42(9): 875-883.

Brooks, Clem and Jeff Manza (2006). "Why Do Welfare States Persist?", *Journal of Politics* 68(4): 816-827.

Calzada, Inés, María Gómez-Garrido, Luis Moreno and Francisco Javier Moreno-Fuentes (2014). "It is not Only About Equality. A Study on the (Other) Values That Ground Attitudes to the Welfare State", *International Journal of Public Opinion Research* 26(2): 178-201.

Carter, John R. and Michael D. Irons (1991). "Are Economists Different, and If So, Why?", *Journal of Economic Perspectives* 5(2): 171-177.

Castles, Francis G. (2007)(ed.). *The Disappearing State? Retrenchment Realities in an Age of Globalisation*. Cheltenham (UK): Edward Elgar.

Chong, Dennis, Jack Citrin and Patricia Conley (2001). "When Self-Interest Matters", *Political Psychology* 22(3): 541-570.

Clausen, Aage R. (1968). "Response Validity: Vote Report", *Public Opinion Quarterly* 32(4): 588-606.

Corstange, Daniel (2009). "Sensitive questions, truthful answers? Modeling the list experiment with LISTIT", *Political Analysis*, 17(1): 45-63.

De Koster, Willem, Peter Achterberg and Jeroen van der Waal (2013). "The new right and the welfare state: The electoral relevance of welfare chauvinism and welfare populism in the Netherlands", *International Political Science Review* 34(1): 3-20.

DePaulo, Bella M., Deborah A. Kashy, Susan E. Kirkendol, Melissa M. Wyer and Jennifer A. Epstein (1996). "Lying in Everyday Life", *Journal of Personality and Social Psychology* 70(5): 979-995.

Dey, Eric L. (1997). "Undergraduate Political Attitudes: Peer Influence in Changing Social Contexts", *Journal of Higher Education* 68(4): 398-413.

Dovidio, John F., Kerry Kawakami, Natalie Smoak and Samuel L. Gaertner (2009). "The Nature of Contemporary Prejudice: Insight from Implicit and Explicit Measures of Attitudes", in R. Petty et al. (eds.), *Attitudes: Insights from the New Implicit Measures*. New York: Psychology Press (165-192).

Droitcour, Judith, Rachel A. Caspar, Michael L. Hubbard, Teresa L. Parsley, Wendy Visscher and Trena M. Ezzati (1991). "The item count technique as a method of indirect questioning: A review of its development and a case study application", in P. Biemer et al. (eds.), *Measurement Errors in Surveys*. New York: Wiley (185-210).

Duff, Brian, Michael J. Hanmer, Won-Ho Park and Ismail K. White (2007). "Good Excuses: Understanding Who Votes with an Improved Turnout Question", *Public Opinion Quarterly* 71(1): 67-90.

Dwight, Stephen A. and John J. Donovan (2003). "Do Warnings Not to Fake Reduce Faking?", *Human Performance* 16(1): 1-23.

Esping-Andersen, Gøsta (1990). *The Three Worlds of Welfare Capitalism*. Cambridge: Polity Press.

Eugster, Beatrix, Rafael Lalive, Andreas Steinhauer and Josef Zweimüller (2011). "The Demand for Social Insurance: Does Culture Matter?", *The Economic Journal* 121: F413-F448.

Feldman, Stanley and Marco R. Steenbergen (2001). "The Humanitarian Foundation of Public Support for Social Welfare", *American Journal of Political Science* 45(3): 658-677.

Fisher, Terri D. (2007). "Sex of Experimenter and Social Norm Effects on Reports of Sexual Behavior in Young Men and Women", *Archives of Sexual Behavior* 36(1): 89-100.

Fiske, Susan T., Monica Lin and Steven L. Neuberg (1999). "The Continuum Model: Ten Years Later", in S. Chaiken and Y. Trope (eds.), *Dual-Process Theories in Social Psychology*. New York: Guilford Press (231-254).

Flavin, Patrick and Michael Keane (2008). "Social Desirability Effects and Opinion toward an African American President". Paper for presentation at the 2008 Annual Meeting of the American Political Science Association, August 28-31.

Freitag, Markus and Adrian Vatter (2006). "Initiatives, referendums, and the tax state", *Journal of European Public Policy* 13(1): 89-112.

Frey, Bruno S. and Stephan Meier (2003). "Are Political Economists Selfish and Indoctrinated? Evidence from a Natural Experiment", *Economic Inquiry* 41(3): 448-462.

Gandal, Neil, Sonia Roccas, Lilach Sagiv and Amy Wrzesniewski (2005). "Personal value priorities of economists", *Human Relations* 58(10): 1227-1252.

Gaxie, Daniel (2007). "Cognitions, auto-habilitation et pouvoirs des 'citoyens'", *Revue Française de Science Politique* 57(6): 737-757.

Gelman, Andrew, Aleks Jakulin, Maria Grazia Pittau and Yu-Sung Su (2008). "A Weakly Informative Default Prior Distribution for Logistic and Other Regression Models", *Annals of Applied Statistics* 2(4): 1360-1383.

Glynn, Adam N. (2010). "What Can We Learn with Statistical Truth Serum? Design and Analysis of the List Experiment". Manuscript, Harvard University.

Glynn, Adam N. (2013). "What Can We Learn with Statistical Truth Serum? Design and Analysis of the List Experiment", *Public Opinion Quarterly* 77(special issue): 159-172.

Gordon, Randall A. (1987). "Social Desirability Bias: A Demonstration and Technique for Its Reduction", *Teaching of Psychology* 14(1): 40-42.

Gray, Mark M. and A. Wuffle (2005). "Vindicating Anthony Downs", *PS: Political Science and Politics* 38(4): 737-740.

Greenwald, Anthony G., T. Andrew Poehlman, Eric Luis Uhlmann and Mahzarin R. Banaji (2009). "Understanding and Using the Implicit Association Test: III. Meta-Analysis of Predictive Validity", *Journal of Personality and Social Psychology* 97(1): 17-41.

Hastie, Brianne (2007). "Higher education and sociopolitical orientation: The role of social influence in the liberalisation of students", *European Journal of Psychology of Education* 22(3): 259-274.

Hayes, Andrew F. (2007). "Exploring the Forms of Self-Censorship: On the Spiral of Silence and the Use of Opinion Expression Avoidance Strategies", *Journal of Communication* 57: 785-802.

Heerwig, Jennifer A. and Brian J. McCabe (2009). "Education and Social Desirability Bias: The Case of a Black Presidential Candidate", *Social Science Quarterly* 90(3): 674-686.

Heinze, Georg (2006). "A comparative investigation of methods for logistic regression with separated or nearly separated data", *Statistics in Medicine* 25: 4216-4226.

Highton, Benjamin (2005). "Self-Reported Versus Proxy-Reported Voter Turnout in the Current Population Survey", *Public Opinion Quarterly* 69(1): 113-123.

Hinnfors, Jonas (2008). "The Puzzling Lack of a Social Dilemma", *Scandinavian Political Studies* 31(1): 69-90.

Holbrook, Allyson L., Melanie C. Green and Jon A. Krosnick (2003). "Telephone versus Face-to-Face Interviewing of National Probability Samples with Long Questionnaires: Comparisons of Respondent Satisficing and Social Desirability Response Bias", *Public Opinion Quarterly* 67(1): 79-125.

Holtgraves, Thomas (2004). "Social Desirability and Self-Reports: Testing Models of Socially Desirable Responding", *Personality and Social Psychology Bulletin* 30(2): 161-172.

Hosmer, David W. and Stanley Lemeshow (2000). *Applied Logistic Regression (2nd ed.)*. Hoboken (NJ): Wiley.

Huckfeldt, Robert and John Sprague (2000). "Political Consequences of Inconsistency: The Accessibility and Stability of Abortion Attitudes", *Political Psychology* 21(1): 57-79.

Huge, Michael E. and Carroll J. Glynn (2013). "Hesitation Blues: Does Minority Opinion Status Lead to Delayed Responses?", *Communication Research* 40(3): 287-307.

Hurtado, Sylvia, Cynthia L. Alvarez, Chelsea Guillermo-Wann, Marcela Cuellar and Lucy Arellano (2012). "A Model for Diverse Learning Environments: The Scholarship on Creating and Assessing Conditions for Student Success", in J. Smart and M. Paulsen (eds.), *Higher Education: Handbook of Theory and Research, vol. 27*. New York: Springer (41-122).

Imai, Kosuke (2011). "Multivariate Regression Analysis for the Item Count Technique", *Journal of the American Statistical Association* 106(494): 407-416.

Jaeger, Mads Meier (2006). "What Makes People Support Public Responsibility for Welfare Provision: Self-interest or Political Ideology? A Longitudinal Approach", *Acta Sociologica* 49(3): 321-338.

Janus, Alexander L. (2007). "The List Experiment as an Unobtrusive Measure of Attitudes Toward Immigration". Paper for presentation at the 102nd Annual Meeting of the American Sociological Association, New York, August 11-14.

Janus, Alexander L. (2010). "The Influence of Social Desirability Pressures on Expressed Immigration Attitudes", *Social Science Quarterly* 91(4): 928-946.

Johnson, Martin (2004). "Timepieces: Components of Survey Question Response Latencies", *Political Psychology* 25(5): 679-702.

Jones, Edward E. and Harold Sigall (1971). "The Bogus Pipeline: A New Paradigm for Measuring Affect and Attitude", *Psychological Bulletin* 76(5): 349-364.

Jones, Elise F. and Jacqueline Darroch Forrest (1992). "Underreporting of Abortion in Surveys of U.S. Women: 1976 to 1988", *Demography* 29(1): 113-126.

Kane, James G., Stephen C. Craig and Kenneth D. Wald (2004). "Religion and Presidential Politics in Florida: A List Experiment", *Social Science Quarterly* 85(2): 281-293.

Kangas, Olli E. (1997). "Self-Interest and the Common Good: The Impact of Norms, Selfishness and Context in Social Policy Opinions", *Journal of Socio-Economics* 26(5): 475-494.

Karp, Jeffrey A. and David Brockington (2005). "Social Desirability and Response Validity: A Comparative Analysis of Overreporting Voter Turnout in Five Countries", *Journal of Politics* 67(3): 825-840.

Kemmelmeier, Markus, Cherry Danielson and Jay Basten (2005). "What's in a Grade? Academic Success and Political Orientation", *Personality and Social Psychology Bulletin* 31(10): 1386-1399.

King, Maryon F. and Gordon C. Bruner (2000). "Social Desirability Bias: A Neglected Aspect of Validity Testing", *Psychology and Marketing* 17(2): 79-103.

Klassen, Deidre, Robijn K. Hornstra and Peter B. Anderson (1975). "Influence of Social Desirability on Symptom and Mood Reporting in a Community Survey", *Journal of Consulting and Clinical Psychology* 43(4): 448-452.

Knoll, Benjamin R. (2013). "Implicit Nativist Attitudes, Social Desirability, and Immigration Policy Preferences", *International Migration Review* 47(1): 132-165.

Korpi, Walter (2001). "Contentious Institutions: An Augmented Rational-Action Analysis of the Origins and Path Dependency of Welfare State Institutions in Western Countries", *Rationality and Society* 13(2): 235-283.

Korpi, Walter (2003). "Welfare-State Regress in Western Europe: Politics, Institutions, Globalization, and Europeanization", *Annual Review of Sociology* 29: 589-609.

Korpi, Walter and Joakim Palme (2003). "New Politics and Class Politics in the Context of Austerity and Globalization: Welfare State Regress in 18 Countries, 1975-95", *American Political Science Review* 97(3): 425-446.

Kreuter, Frauke, Stanley Presser and Roger Tourangeau (2008). "Social Desirability Bias in CATI, IVR, and Web Surveys: The Effects of Mode and Question Sensitivity", *Public Opinion Quarterly* 72(5): 847-865.

Kriesi, Hanspeter and Alexander H. Trechsel (2008). *The Politics of Switzerland: Continuity and Change in a Consensus Democracy*. Cambridge: Cambridge University Press.

Krosnick, Jon A. (1999). "Maximizing Questionnaire Quality", in J. Robinson, P. Shaver & L. Wrightsman (eds.), *Measures of Political Attitudes*. San Diego: Academic Press (37-57).

Krosnick, Jon A. and Stanley Presser (2010). "Question and Questionnaire Design", in J. Wright & P. Marsden (eds.), *Handbook of Survey Research (2nd ed.)*. West Yorkshire (UK): Emerald Group (263-313).

Krysan, Maria (1998). "Privacy and the Expression of White Racial Attitudes: A Comparison Across Three Contexts", *Public Opinion Quarterly* 62(4): 506-544.

Krysan, Maria and Mick P. Couper (2003). "Race in the Live and the Virtual Interview: Racial Deference, Social Desirability, and Activation Effects in Attitude Surveys", *Social Psychology Quarterly* 66(4): 364-383.

Krysan, Maria and Mick P. Couper (2006). "Race of Interviewer Effects: What Happens on the Web?", *International Journal of Internet Science* 1(1): 17-28.

Kuklinski, James H., Michael D. Cobb and Martin Gilens (1997). "Racial Attitudes and the 'New South'", *Journal of Politics* 59(2): 323-349.

Kulin, Joakim and Stefan Svallfors (2013). "Class, Values, and Attitudes Towards Redistribution: A European Comparison", *European Sociological Review* 29(2): 155-167.

Kumlin, Staffan (2004). *The Personal and the Political: How Personal Welfare State Experiences Affect Political Trust and Ideology*. New York: Palgrave Macmillan.

Kumlin, Staffan (2007). "The Welfare State: Values, Policy Preferences, and Performance Evaluations", in R. Dalton & H. Klingemann (eds.), *The Oxford Handbook of Political Behavior*. Oxford: Oxford University Press (362-382).

Lax, Jeffrey R., Justin H. Phillips and Alissa F. Stollwerk (2014). "Are Survey Respondents Lying About their Support for Same-Sex Marriage? Lessons from a Recent List Experiment". Unpublished manuscript.

Levine, Mark, Amy Prosser, David Evans and Stephen Reicher (2005). "Identity and Emergency Intervention: How Social Group Membership and Inclusiveness of Group Boundaries Shape Helping Behavior", *Personality and Social Psychology Bulletin* 31(4): 443-453.

Linos, Katerina and Martin West (2003). "Self-interest, Social Beliefs, and Attitudes to Redistribution: Re-addressing the Issue of Cross-national Variation", *European Sociological Review* 19(4): 393-409.

Liu, Amy, Sylvia Ruiz, Linda DeAngelo and John Pryor (2009). "Findings from the 2008 Administration of the College Senior Survey (CSS): National Aggregates", Higher Education Research Institute, University of California, Los Angeles.

Lord, Charles G. and Mark R. Lepper (1999). "Attitude Representation Theory", in M. Zanna (ed.), *Advances in Experimental Social Psychology, Vol. 31*. San Diego (CA): Academic Press (265-343).

Lord, Charles G., René M. Paulson, Tiffiny L. Sia, Jennifer C. Thomas and Mark R. Lepper (2004). "Houses Built on Sand: Effects of Exemplar Stability on Susceptibility to Attitude Change", *Journal of Personality and Social Psychology* 87(6): 733-749.

Manley, Gregory G., Craig J. Russell and M. Ronald Buckley (2001). "Self-Enhancing in Perceptions of Behaving Unethically", *Journal of Education for Business* 77(1): 21-27.

Manow, Philip (2004). "The Good, the Bad, and the Ugly: Esping-Andersen's Regime Typology and the Religious Roots of the Western Welfare State", MPIfG Working Paper No. 04/3, Max Planck Institute for the Study of Societies, Cologne, Germany.

Marquis, Lionel, Hans-Peter Schaub and Marlène Gerber (2011). "The Fairness of Media Coverage in Question: An Analysis of Referendum Campaigns on Welfare State Issues in Switzerland", *Swiss Political Science Review* 17(2): 128-163.

Mattei, Franco (1998). "Winning at the Polls and in the Polls: the Incumbency Advantage in Surveys of U.S. House Voters", *Electoral Studies* 17(4): 443-461.

Matthes, Jörg, Kimberly Rios Morrison and Christian Schemer (2010). "A Spiral of Silence for Some: Attitude Certainty and the Expression of Political Minority Opinions", *Communication Research* 37(6): 774-800.

Matthews J. Scott and Lynda Erickson (2008). "Welfare State Structures and the Structure of Welfare State Support: Attitudes Towards Social Spending in Canada, 1993-2000", *European Journal of Political Research* 47(4): 411-435.

Mau, Steffen (2003). *The Moral Economy of Welfare States: Britain and Germany Compared*. London: Routledge.

Mesmer-Magnus, Jessica, Chockalingam Viswesvaran, Satish Deshpande and Jacob Joseph (2006). "Social desirability: the role of over-claiming, self-esteem, and emotional intelligence", *Psychology Science* 48(3): 336-356.

Meston, Cindy M., Julia R. Heiman, Paul D. Trapnell and Delroy L. Paulhus (1998). "Socially Desirable Responding and Sexuality Self-Reports", *Journal of Sex Research* 35(2): 148-157.

Milfont, Taciano L. (2009). "The effects of social desirability on self-reported environmental attitudes and ecological behaviour", *Environmentalist* 29: 263-269.

Miller, Dale T. (2001). "The Norm of Self-Interest", in M. Schwartz & H. Harris (eds.), *Research in Ethical Issues in Organizations, vol. 3*. Greenwich (CT): JAI Press (193-210).

Nederhof, Anton J. (1985). "Methods of coping with social desirability bias: a review", *European Journal of Social Psychology* 15: 263-280.

Neuwirth, Kurt, Edward Frederick and Charles Mayo (2007). "The Spiral of Silence and Fear of Isolation", *Journal of Communication* 57: 450-468.

Nolan, Jessica M., P. Wesley Schultz, Robert B. Cialdini, Noah J. Goldstein and Vladas Griskevicius (2008). "Normative Social Influence is Underdetected", *Personality and Social Psychology Bulletin* 34(7): 913-923.

Nosek, Brian A. (2005). "Moderators of the Relationship Between Implicit and Explicit Evaluation", *Journal of Experimental Psychology: General* 134(4): 565-584.

Obinger, Herbert, Klaus Armingeon, Giuliano Bonoli and Fabio Bertozzi (2005). "Switzerland: The Marriage of Direct Democracy and Federalism", in H. Obinger et al. (eds.), *Federalism and the Welfare State: New World and European Experiences*. New York: Cambridge University Press (263-304).

Ones, Deniz S. and Chockalingam Viswesvaran (1998). "The Effects of Social Desirability and Faking on Personality and Integrity Assessment for Personnel Selection", *Human Performance* 11(2-3): 245-269.

Pascarella, Ernest T. and Patrick T. Terenzini (2005). *How College Affects Students. Volume 2: A Third Decade of Research*. San Francisco: Jossey-Bass.

Paulhus, Delroy L. (1984). "Two-component Models of Socially Desirable Responding", *Journal of Personality and Social Psychology* 46(3): 598-609.

Paulhus, Delroy L. (2002). "Socially Desirable Responding: The Evolution of a Construct", in H. Braun, D. Jackson & D. Wiley (eds.), *The Role of Constructs in Psychological and Educational Measurement*. Mahwah (NJ): Lawrence Erlbaum (49-69).

Paulhus, Delroy L. and Douglas B. Reid (1991). "Enhancement and Denial in Socially Desirable Responding", *Journal of Personality and Social Psychology* 60(2): 307-317.

Pierson, Paul (1994). *Dismantling the Welfare State? Reagan, Thatcher, and the Politics of Retrenchment*. Cambridge: Cambridge University Press.

Prasad, Monica (2006). *The Politics of Free Markets: The Rise of Neoliberal Economic Policies in Britain, France, Germany, and the United States*. Chicago: University of Chicago Press.

Presser, Stanley and Michael Traugott (1992). "Little White Lies and Social Science Models: Correlated Response Errors in a Panel Study of Voting", *Public Opinion Quarterly* 56(1): 77-86.

Presser, Stanley and Linda Stinson (1998). "Data Collection Mode and Social Desirability Bias in Self-Reported Religious Attendance", *American Sociological Review* 63(1): 137-145.

Pryor, John H., Kevin Eagan, Laura Palucki Blake, Sylvia Hurtado, Jennifer Berdan and Matthew H. Case (2012). *The American Freshman: National Norms Fall 2012 (Expanded Edition)*. Los Angeles: Higher Education Research Institute, UCLA.

Reicher, Steve and Nick Hopkins (1996). "Seeking influence through characterizing self-categories: An analysis of anti-abortionist rhetoric", *British Journal of Social Psychology* 35: 297-311.

Richman, Wendy L., Sara Kiesler, Suzanne Weisband and Fritz Drasgow (1999). "A Meta-Analytic Study of Social Desirability Distortion in Computer-Administered Questionnaires, Traditional Questionnaires, and Interviews", *Journal of Applied Psychology* 84(5): 754-775.

Roosma, Femke, John Gelissen and Wim van Oorschot (2013). "The Multidimensionality of Welfare State Attitudes: A European Cross-National Study", *Social Indicators Research* 113(1): 235-255.

Ross, Fiona (2000). "Interests and Choice in the 'Not Quite so New' Politics of Welfare", *West European Politics* 23(2): 11-34.

Schaeffer, Nora Cate (1999). "Asking Questions About Threatening Topics: A Selective Overview", in A. Stone et al. (eds.), *The Science of Self-Report: Implications for Research and Practice*. Mahwah (NJ): Lawrence Erlbaum (105-121).

Schwartz, Barry (2012). "Crowding Out Morality: How the Ideology of Self-Interest Can Be Self-Fulfilling", in J. Hanson (ed.), *Ideology, Psychology, and Law*. New York: Oxford University Press (160-184).

Schwarz, Norbert (2007). "Attitude Construction: Evaluation in Context", *Social Cognition* 25(5): 638-656.

Schwarz, Norbert and Gerd Bohner (2001). "The Construction of Attitudes", in A. Tesser & N. Schwarz (eds.), *Blackwell Handbook of Social Psychology: Intraindividual Processes*. Malden (MA): Blackwell (436-457).

Smith, Eliot R. and Frederica R. Conrey (2007). "Mental Representations Are States, Not Things", in B. Wittenbrink & N. Schwarz (eds.), *Implicit Measures of Attitudes*. New York: Guilford Press (247-264).

Sniderman, Paul M. and Edward G. Carmines (1997). *Reaching Beyond Race*. Cambridge (MA): Harvard University Press.

Sniderman, Paul M. and Louk Hagendoorn (2007). *When Ways of Life Collide: Multiculturalism and Its Discontents in the Netherlands*. Princeton: Princeton University Press.

Stocké, Volker and Tobias Stark (2007). "Political Involvement and Memory Failure as Interdependent Determinants of Vote Overreporting", *Applied Cognitive Psychology* 21: 239-257.

Streb, Matthew J., Barbara Burrell, Brian Frederick and Michael A. Genovese (2008). "Social Desirability Effects and Support for a Female American President", *Public Opinion Quarterly* 72(1): 76-89.

Svallfors, Stefan (1997). "Worlds of Welfare and Attitudes to Redistribution: A Comparison of Eight Western Nations", *European Sociological Review* 13(3): 283-304.

Taylor-Gooby, Peter (2001). "Sustaining state welfare in hard times: who will foot the bill?", *Journal of European Social Policy* 11(2): 133-147.

Tourangeau, Roger and Kenneth A. Rasinski (1988). "Cognitive Processes Underlying Context Effects in Attitude Measurement", *Psychological Bulletin* 103(3): 299-314.

Tourangeau, Roger and Ting Yan (2007). "Sensitive Questions in Surveys", *Psychological Bulletin* 133(5): 859-883.

Tourangeau, Roger, Lance J. Rips and Kenneth Rasinski (2000). *The Psychology of Survey Response*. Cambridge: Cambridge University Press.

Trampusch, Christine and André Mach (2011)(eds.). *Switzerland in Europe: Continuity and Change in the Swiss Political Economy*. London: Routledge.

Tsuchiya, Takahiro, Yoko Hirai and Shigeru Ono (2007). "A Study of the Properties of the Item Count Technique", *Public Opinion Quarterly* 71(2): 253-272.

Umesh, U. N. and Robert A. Peterson (1991). "A Critical Evaluation of the Randomized Response Method: Applications, Validation, and Research Agenda", *Sociological Methods & Research* 20(1): 104-138.

Van der Pligt, Joop and Nanne K. de Vries (1998). "Belief Importance in Expectancy-Value Models of Attitudes", *Journal of Applied Social Psychology* 28(15): 1339-1354.

Van Harreveld, Frenk and Joop van der Pligt (2004). "Attitudes as stable and transparent constructions", *Journal of Experimental Social Psychology* 40: 666-674.

Van Kersbergen, Kees (2000). "The declining resistance of welfare states to change?", in S. Kuhnle (ed.), *Survival of the European Welfare State*. London: Routledge (19-36).

Von Hippel, William and Robert Trivers (2011). "The evolution and psychology of self-deception [followed by Open Peer Commentary]", *Behavioral and Brain Sciences* 34(1): 1-56.

Wänke, Michaela, Herbert Bless and Barbara Biller (1996). "Subjective Experience Versus Content of Information in the Construction of Attitude Judgments", *Personality and Social Psychology Bulletin* 22(11): 1105-1113.

Wilson, Timothy D. and Sara D. Hodges (1992). "Attitudes as Temporary Constructions", in L. Martin & A. Tesser (eds.), *The Construction of Social Judgments*. Hillsdale (NJ): Lawrence Erlbaum (37-65).

Wood, Wendy (2000). "Attitude Change: Persuasion and Social Influence", *Annual Review of Psychology* 51: 539-570.

Wright, Gerald C. (1993). "Errors in Measuring Vote Choice in the National Election Studies, 1952-88", *American Journal of Political Science* 37(1): 291-316.

Yzerbyt, Vincent and Toon Kuppens (2009). "Group-based emotions", in S. Otten et al. (eds.), *Intergroup Relations: The Role of Motivation and Emotion*. London: Psychology Press (143-161).

Zaller, John R. and Stanley Feldman (1992). "A Simple Theory of the Survey Response: Answering Questions versus Revealing Preferences", *American Journal of Political Science* 36(3): 579-616.

Zeglovits, Eva and Sylvia Kritzinger (2014). "New Attempts to Reduce Overreporting of Voter Turnout and Their Effects", *International Journal of Public Opinion Research* 26(2): 224-234.

Zigerell, Lawrence J. (2011). "You Wouldn't Like Me When I'm Angry: List Experiment Misreporting", *Social Science Quarterly* 92(2): 552-562.

Zorn, Christopher (2005). "A Solution to Separation in Binary Response Models", *Political Analysis* 13(2): 157-170.

Dans la collection « Cahiers de l'IEPHI » /« Travaux de Science Politique »

Disponibles sur www.unil.ch/iephi

68. **Dimitri Courant** (2017). Thinking Sortition. Modes of selection, deliberative frameworks and democratic principles.
67. **Oscar Mazzoleni, Carolina Rossi, Andrea Pilotti et Virginie Debons** (2017). Le elezioni cantonali ticinesi del 2015. Partecipazione, orientamento, temi e personalizzazione del voto.
66. **Maxime Filliau, Aziz Haltiti, Anna Herczeg, Loeva La Ragione, Régis Marchon, Baudouin Noez, Milinda Wannakula Aratchilage, Yuzhou Wu, Jan-Erik Refle et Lionel Marquis** (2016). Immigration, sécurité et comportement électoral Les Européens face aux crises économique, migratoire et sécuritaire.
65. **Nicholas Pohl** (2016). Quand les hiérarchies reviennent par la fenêtre… La démocratie au sein d'un syndicat anarcho-syndicaliste.
64. **Philippe Gottraux et Cécile Péchu** (2016). Sous le populisme, la pluralité. Anti-élitisme et rapports au leader chez les militants de l'Union Démocratique du Centre en Suisse.
63. **Michele Bee** (2015). Opening One's Self Up The Historical Result of Bettering One's Condition According to Adam Smith.
62. **Conor Cradden et Jean-Christophe Graz** (2015). Transnational private authority, regulatory space and workers' collective competences: Bringing local contexts and worker agency back in.
61. **Andrea Pilotti et Oscar Mazzoleni** (2014). Il voto ticinese sull'iniziativa 'contro l'immigrazione di massa' del 9 febbraio 2014.
60. **Michele Bee** (2014). The Love of One's Self. The *Adam Smith Problem* Explained to Myself.
59. **Marion Beetschen** (2014). Les experts académiques dans les commissions extraparlementaires helvétiques au XXe siècle.
58. **Rahel Kunz** (2013). Partnership in International Migration Governance.
57. **Gilles Descloux** (2013). Déviance et stratégies de conservation de l'estime de soi chez les jeunes de catégorie populaire en quête d'insertion professionnelle.
56. **Pierre Eichenberger, Sébastien Guex, Cédric Humair et André Mach** (2013). Les organisations patronales suisses : Bilan historiographique et perspectives de recherche.
55. **Andrea Plata, Mattia Pacella et Oscar Mazzoleni** (2012). Verso il parlamento. Candidati, partiti e campagna elettorale nelle elezioni cantonali ticinesi del 2011.
54. **Jean-Christophe Graz et Patrick Csikos** (2012). Satellizing Galileo? Non-state authority and interoperability standards in the European Global Navigation Satellite System.
53. **Claire Dupuy et Julie Pollard** (2012). Retour sur une source écrite centrale pour l'analyse des politiques publiques.
52. **Jean-Marie Chenou** (2012). Élitisme dans la politique mondiale. Le rôle des acteurs transnationaux dans les nouvelles formes de gouvernance.
51. **Oscar Mazzoleni, Mattia Pacella et Andrea Plata** (2011). Le elezioni cantonali ticinesi del 2011. Partecipazione e orientamento di voto fra lealtà, defezione, protesta.
50. **Olivier Fillieule** (2011). Disengagement process from radical organizations. What is so different when it comes to exclusive groups?
49. **Olivier Fillieule et Philippe Blanchard** (2011). Fighting Together. Assessing Continuity and Change in Social Movement Organizations Through the Study of Constituencies' Heterogeneity.
48. **Jean-Christophe Graz et Eva Hartmann** (2010). Global regulation of the knowledge-based economy: The rise of standards in educational services.
47. **Jean-Christophe Graz et Nafy Niang** (2010). Standards and services offshoring in India: Moving beyond industry and institutional specificities.
46. **Maude Gex** (2010). Le divin profané par Da Vine ? La Contestation du brevet sur

l'ayahuasca et les débats autour de la biopiraterie.

45. **Mattia Pacella** (2010). Qui gouverne le Tessin ? Les Elites politiques cantonales tessinoises : l'évolution du profil sociologique des Conseillers d'Etat de 1905 à nos jours.
44. **Aline Mugny** (2010). Analyse du processus de réétatisation du secteur des eaux urbaines dans l'aire métropolitaine du Buenos Aires (1993-2009).
43. **Nicolas Rieder** (2009). La diffusion scientifique dans les Expositions universelles. Les sciences anthropologiques à l'Exposition universelle de Paris en 1889 et à la World's Columbian Exposition de Chicago en 1893.
42. **Philip Balsiger, Yassin Boughaba et Gwendoline Véniat** (2009). Conflit du travail et investissements militants. Les grèves des employés.es de Swissmetal à Reconvilier (2004-2006).
41. **Dietmar Braun** (2009). Federalism in South Africa – Can it work?
40. **Romain Bertrand** (2009). Habermas au Bengale, ou comment "provincialiser l'Europe" avec Dipesh Chakrabarty.
39. **Oscar Mazzoleni et Hervé Rayner** (2008). Une coalition gouvernementale durable. Emergence, institutionnalisation et crise de la «formule magique» en Suisse (1959-2003).
38. **Andrin Hauri** (2008). Les conséquences de la barrière de séparation entre Israël et la Cisjordanie à court et à long terme. Les perspectives pour la fondation d'un État palestinien.
37. **André Mach et Andrea Pilotti** (2008). Professionnalisation et changements de profils des parlementaires suisses au cours de la période récente (1980-2000)
36. **Olivier Fillieule** (2008). Travail militant, action collective et rapports de genre.
35. **Olivier Fillieule et Philippe Blanchard** (2008). INdividual SUrvey in RAllies (INSURA). A New Tool for Exploring Transnational Activism.
34. **Johann Dupuis** (2008). Analyse politique des conditions de succès et d'échec des marchés de droits d'émissions.
33. **Dietmar Braun et Björn Uhlmann** (2007). Ideas and Power in Swiss Health Care Party Politics.
32. **Fabrizio Gilardi, Katharina Füglister et Stéphane Luyet** (2007). Learning from Others: The Diffusion of Hospital Financing Reforms in OECD Countries.
31. **Natalia Gerodetti** (2007). Enduring Legacies – Intersecting Discourses in the Context of Eugenics.
30. **Véronique Mottier** (2007). Meaning, Identity, Power : Metaphors and Discourse Analysis.
29. **Olivier Fillieule** (2007). On n'y voit rien ! Le recours aux sources de presse pour l'analyse des mobilisations protestataires.
28. **Kathrin Daepp** (2006). La re-régulation sélective de la finance internationale : l'initiative sur les pays ou territoires non-coopératifs du GAFI.
27. **Martino Maggetti** (2006). Assessing the De Facto Independence of Regulatory Agencies. The Case of the Swiss Federal Banking Commission in the 1990?
26. **Dietmar Braun** (2006). Modernising Federalism. Towards Convergence in the Organisation of Intergovernmental Relations?
25. **Fabrizio Gilardi** (2006). The Same, but Different. Central Banks, Regulatory Agencies, and the Politics of Delegation to Independent Authorities.
24. **Cécile Péchu** (2006). Entre résistance et contestation. La genèse du squat comme mode d'action.
23. **Sarah Gotheil** (2005) ACCOBAMS. Les pays de mer noire, méditerranée et zone atlantique adjacente s'unissent pour la protection des cétacés.
22. **Gerhard Schnyder, Martin Lüpold, André Mach et Thomas David** (2005) The Rise and Decline of the Swiss Company Network during the 20th Century.
21. **Aurélien Buffat** (2005) La mise en oeuvre de la réforme de l'armée suisse « Armée XXI » : les changements vécus à l'interne. Etude de cas des militaires professionnels des écoles, hôpital de la place d'armes de Moudon.
20. **Léonard Rey** (2005) La construction du marché intérieur suisse à la lumière de l'expérience européenne : centralité et intervention contrastée des Hautes Cours dans les processus d'intégration économique.
19. **Patrick Csikos** (2005) Emergence d'une nouvelle gouvernance internationale privée/publique : les cas des agences de notation financière et des normes comptables.
18. **Michael Voegtli** (2004) Entre paternalisme et Etat social. Le cas de la fabrique de chocolat Suchard (1870-1940).

17. **Luca Chinotti** (2004) Les effets de l'Accord sur l'Agriculture de l'Uruguay Round sur les pays en développement. Une entrave ou une opportunité pour le développement?.
16. **Alexandre Afonso** (2004) Internationalisation, économie et politique migratoire dans la Suisse des années 1990.
15. **Nicolas Freymond** (2003) La question des institutions dans la science politique contemporaine: l'exemple du néo-institutionnalisme.
14. **Yves Steiner** (2003) Le coût réel de l'indépendance de la banque centrale: économie politique comparée de la Deutsche Bundesbank et de la Banque du Japon dans les années soixante-dix.
13. **Dietmar Braun** (1999) Toward a heuristic framework of the Territorial division of Power in comparative public Policy research.
12. **Dietmar Braun** (1998) Bringing State structures back in: The Significance of Political Arena's in Political Decision-making.
11. **Daniel Kübler** (1996) Neighbourhood conflicts and dialogues.
10. **André Mach** (1995) Représentation des intérêts et capacité d'adaptation de l'économie suisse.
9. **Yves Sancey** (1995) Le Gentlemen's agreement de 1927. Lutte autour de la (non-)politisation de l'exportation du capital.
8. **Daniel Kübler** (1993) L'Etat face à la toxicomanie, action publique et contrat social.
7. **Jean-Philippe Leresche** (1993) Les transformations du pouvoir local en Suisse.
6. **Bernard Voutat** (1993) Les minorités territoriales, quelques considérations théoriques.
5. **Bernard Voutat** (1992) Les origines sociales de la "question jurassienne".
4. **Daniel-Louis Seiler** (1991) Le cas des partis politiques dans les nouvelles démocraties de l'Est européen.
3. **Silvia Kobi** (1991) Les "Neinsager" dans le processus référendaire suisse: des variations sur un thème mythique.
2. **Ioannis Papadopoulos** (1991) La Suisse: un "Sonderfall" pour la théorie politique?

   Nadia Spang (1991) Scholarly View of Japan throughout the XXth Century.

   Shahrokh Vaziri (1990) Eléments de réflexion sur les partis politiques et références idéologiques dans le Tiers Monde.