The Politics of Trade-offs: Studying the Dynamics of Welfare State Reform with Conjoint Experiments*†

Silja Häusermann‡, Thomas Kurer§, Denise Traber¶

Accepted for publication in *Comparative Political Studies*

Abstract

Welfare state reform in times of austerity is notoriously difficult because most citizens oppose retrenchment of social benefits. Governments thus tend to combine cutbacks with selective benefit expansions, thereby creating trade-offs: in order to secure new advantages, citizens must accept painful cutbacks. Prior research has been unable to assess the effectiveness of compensating components in restrictive welfare reforms. We provide novel evidence on feasible reform strategies by applying conjoint survey analysis to a highly realistic direct democratic setting of multidimensional welfare state reform. Drawing on an original survey of Swiss citizens' attitudes toward comprehensive pension reform, we empirically demonstrate that built-in trade-offs strongly enhance the prospects of restrictive welfare reforms. Our findings indicate that agency matters: governments and policymakers can and must grant the right compensations to the relevant opposition groups in order to overcome institutional inertia.

*Earlier versions of this article have been presented at the ECPR Joint Sessions 2016 in Pisa, the Social Policy Preferences Workshop in Duke 2016, at the Swiss Political Science Conference 2017 in St. Gallen, the conjoint workshop 2017 at the University of Zurich, the Nuffield College Politics Seminar in April 2017, and the European Political Science Association annual meeting in Milan 2017. We thank the participants of these conferences for their comments. We are particularly grateful to Tarik Abou-Chadi, Cameron Ballard-Rosa, Aina Gallego, Jane Gingrich, Dominik Hangartner, Yotam Margalit, Erik Neimanns, Kenneth Scheve, and Teppei Yamamoto for helpful suggestions and comments.

†This work was supported by Swiss National Science Foundation (Schweizerischer Nationalfonds zur Förderung der Wissenschaftlichen Forschung) [grant number 100017_159341 / 1].

‡Department of Political Science, University of Zurich; silja.haeusermann@ipz.uzh.ch.

§Department of Political Science, University of Zurich; kurer@ipz.uzh.ch.

¶Department of Political Science, University of Lucerne; denise.traber@unilu.ch.

1 Introduction

The modern welfare state is one of the major social and political achievements of the twentieth-century post-war era in the developed OECD world. It has supported democratic stability and has allowed for the shielding of most citizens throughout Europe from the main risks of income loss and poverty, such as sickness, old age, and unemployment. At the beginning of the twenty-first century, however, European welfare states have come under heavy political pressure because demographic and economic structural change call for financial consolidation or even retrenchment of welfare benefits (Pierson, 2001b; Huber & Stephens, 2001). At the same time, new social risks and economic grievances have increased social demands on the welfare state (Pierson, 2001b; Bonoli, 2005). Beyond the need for financial stabilization, the sustainability of welfare states also depends on their capacity for adaptation. For these reasons, understanding the reform capacity of welfare states has become a key topic in comparative politics (Esping-Andersen, 1999; Pierson, 2001b; Huber & Stephens, 2001; Hacker, 2004; Kuhnle, 2002; Immergut, Anderson & Schulze, 2007; Brooks & Manza, 2008; Palier, 2010; Häusermann, 2010; Rehm, Hacker & Schlesinger, 2012; Hemerijck, 2013; Gingrich, 2014; Huber & Stephens, 2015; Lindvall, 2017). All these studies show that—despite strong functional pressure for welfare reform—the financial consolidation of major welfare programs and their adaptation to new social demands are highly conflictual politically since institutional feedback processes create endogenous stabilizers for existing welfare state programs (Pierson, 2000). These endogenous stabilizers entail both functional and political costs of policy change for elites and voters (Pierson, 2001a).

Hence, key institutional stabilization mechanisms of welfare states are based on citizens' preferences and public opinion: reforms are difficult either because citizens protest against cutbacks (in the electoral or non-electoral arena), or because elites fear the repercussions of reforms and use different blame-avoidance strategies (Vis, 2016). Several studies (Giger & Nelson, 2011; Cox, 2001) show that not all retrenchment is equally unpopular. Particularly where risks are concentrated (e.g., unemployment, see Rehm et al. (2012)), public opinion may support rather than oppose cost containment. However,

public opinion is clearly opposed to cutbacks when it comes to "life-cycle risks" (Jensen, 2012) such as health, old age, and education. Therefore, the key question for welfare state research in the twenty-first century has become under what conditions existing social benefits can be cut back in the context of public opinion opposing such cutbacks.

The main answer to this question in the welfare state literature is that significant welfare reform is possible when governments manage to diffuse resistance by *compensating* (potential) opponents. A great number of empirical studies have established such political exchange as a key mechanism of current welfare reform politics, especially in power-sharing political systems (Levy, 1999; Bonoli, 2000; Pierson, 2001b; Natali & Rhodes, 2004; Häusermann, 2010; Huber & Stephens, 2015; Knotz & Lindvall, 2015; Lindvall, 2017). Their findings imply that unpopular reforms need to be counterbalanced to be viable politically. If a reform contains both contested as well as popular elements, individuals and collective actors are confronted with a trade-off: while they reject the reform based on the cutbacks, they may be interested in the reform being adopted based on the compensating aspects. Eventually, citizens may support the overall package despite their opposition to certain parts of the reform. The aggregate support for the reform will depend on the *relative importance* individuals or social groups attribute to the different elements of the reform. This argument has prominently brought agency back into a literature that had become predominantly focused on mechanisms of institutional path-dependency and stability: if it is true that governments can skillfully "design" reforms in a way that secures sufficient popular support, then the room for politics and choice is reestablished.

However, despite the theoretical prominence of the compensation argument, prior research has to date been unable to assess the impact of compensating expansive reform elements on attitudes toward restrictive welfare reforms because such reforms are multi-dimensional. In other words, we do not have empirical knowledge about the effectiveness of different kinds of compensation strategies. Who needs to be compensated and which groups are (most) receptive to compensations? The reason for this lack of research on the micro-foundations of welfare reform is that standard survey data does not provide

information on preferences regarding multidimensional policy reforms. As a consequence, all we can do is infer the effectiveness of compensations ex post facto and indirectly (e.g., via electoral outcomes (Giger & Nelson, 2013)) after unpopular reform elements have been implemented. However, this kind of backward induction implies that we can never test the mechanism of compensation itself, and neither can we empirically identify the relative effectiveness of different compensation strategies.

In this paper, we introduce experimental conjoint survey analysis to study the political dynamics of compensation in welfare reform. Conjoint analysis prompts respondents to choose between different *policy packages* rather than simply asking about support for different specific measures. The packages contain diverse reform elements, both elements of retrenchment and compensation, and therefore allow us to study the respective contribution of each individual reform element to overall support in the context of multidimensional welfare reform.

Our empirical case is the most recent pension reform in Switzerland, a reform process that entails different compensation strategies in the one social policy field—old-age pensions—where benefit retrenchment and financial consolidation are the least popular (Jensen, 2012). Switzerland provides an ideal context for studying social policy preferences because citizens have the possibility of asking for a vote on major welfare reforms in direct democratic referenda. Hence, we are able to study preferences for welfare retrenchment and compensation in a highly realistic setting.

We find that, while pension retrenchment quite dramatically reduces support for a reform, compensation via recalibrating (i.e., the adaptation of social policies to changed family patterns and employment biographies) and targeting elements (i.e., retrenchment is combined with benefit expansions for lower-income beneficiaries) can counterbalance the cost of cutbacks. Generally, retrenching reform packages that include compensation generate more support than reform proposals including very weak or no compensation. This finding holds for all respondents, even those who do not perceive a very high reform pressure. With regard to specific social groups, our results show that targeted pension improvements and recalibrating reform elements effectively increase support among low-

income earners and women, their main beneficiaries. Moreover, the findings also point to a stronger effect of ideology (partisanship) on preferences for specific compensation strategies compared to socio-economic determinants. Overall, we conclude that compensating elements, especially recalibration, effectively enhance the prospects of restrictive welfare reforms.

Our findings are relevant beyond the Swiss context because even in countries where reforms are not voted on at the polls, governments and political parties are sensitive to public opinion, and citizens can challenge and even veto unpopular and unbalanced reforms via protest, strikes, and other forms of direct involvement. And while the direct democratic scenario obviously introduces a relatively high degree of factual detail in the reform proposals that respondents were asked to evaluate, we made sure to include in our survey only reform issues that were highly politicized and visible in the public debate (i.e. which would most likely also raise debates in other contexts).

2 Theory: The Politics of Trade-Offs in Welfare State Reform

Compensation and political exchange play a key role in today's welfare reform capacity. This argument is largely uncontested in the welfare state literature (Bonoli & Natali, 2012; Huber & Stephens, 2015; Knotz & Lindvall, 2015; Lindvall, 2017). Of course, there are specific institutional and political circumstances in which even harsh cutbacks can be imposed by governments, for example when power is undivided and governments do not fear the electoral backlash (Kitschelt, 2001) or when retrenchment affects either a small group that bears concentrated risks (Jensen, 2012) or social groups that are considered "undeserving" (Van Oorschot, 2006). But when it comes to the established central pillars of the welfare state on which the financial stability of the entire regime ultimately depends—and old age pensions are the prime example of those—such unilateral, uncompensated cutbacks have turned out to be politically unviable (Pierson, 2001b; Häusermann, 2010; Vail, 2010; Jensen, 2012). The main reason uncompensated pension cutbacks are unvi-

able is that existing benefits have very strong support among a broad majority of the citizens, and thus the defenders of existing benefit levels and benefit structures have political visibility and clout, both in the public debates and in organized politics.[1] In a recent study, Busemeyer and Garritzmann (2017) confirmed that when confronted with trade-offs between several desirable goals, people were particularly unwilling to accept cutbacks in pension benefits. In other words, compensating the opponents of a reform is crucial for a government that wants to cut back pension benefits.

The unpopularity of pension cutbacks is strongly reflected in our survey. About 60 percent of the respondents claim that lower benefits are "not acceptable."[2] Refusal is high (above 50 percent) throughout the sample, even among upper-income classes and across the political spectrum. However, some groups stand out in their opposition. On the one hand, more vulnerable socio-structural categories (lower-income earners and women) are particularly opposed to lower benefits.[3] On the other hand, party affiliation matters: voters of the far right and the left are the main opponents of cutbacks. The opposition among left voters is generally a more serious threat to cost-saving reforms because the far right in Switzerland promotes lower taxes and market liberalism (Bornschier, 2015) and is therefore very unlikely to challenge pension retrenchment despite the more skeptical stance of such voters. The left, in contrast, has a history of challenging any retrenchment in direct democratic referenda.[4]

Existing research has conceptualized and observed different kinds of compensation, which can be grouped into three strategies, based on the specific sociopolitical groups

[1] In the ISSP 2006 comparative survey data, less than 10% of respondents in all European countries supported the statement that the "state should spend less or much less" on old age pensions.

[2] Respondents were asked how they would prefer to consolidate pension finances: (a) by balancing retrenchment and increased revenues, (b) primarily through cutbacks, since higher taxes are not acceptable (c) primarily through higher revenues since lower benefits are not acceptable. Over 60 percent of all respondents chose the answer (c).

[3] Switzerland has a typical continental (Palier, 2010) pension system based on earnings-related social insurance. This overall architecture benefits those with continuous employment biographies and those in the middle- to upper-income classes, which is why old-age poverty risks are significantly higher for lower-income classes as well as women.

[4] Two referenda against pension cutbacks in 2003 and 2010 were launched by the left and successfully blocked reforms. In the fall of 2017, the radical wing of the Swiss left also fought against the current reform in the referendum campaign because it was fiscally conservative thereby joining in an "unholy alliance" with right-wing liberal actors who thought that compensation had gone too far. Together, this alliance against the reform defied the proposal in the referendum with 53 percent of the votes.

that are compensated. First, opponents can be compensated via *targeting* i.e., retrenchment is combined with benefit expansions for lower-income beneficiaries. The negative effects of cutbacks are thereby mitigated for the most vulnerable, usually through the strengthening of means-tested aspects of the social policy schemes (for examples of pension and labor market reforms in Continental and Southern Europe, see Rhodes, 2001; Knotz & Lindvall, 2015). Targeting is supposed to foster support for the reform among the direct beneficiaries of targeted expansions (i.e., the lower-income voters) as well as among left-wing voters who are opposed to greater inequality.

The second compensation strategy is *recalibration*. Recalibration links benefit cutbacks to welfare state "updating" (Pierson, 2001b), i.e., the adaptation of social policies to changed family patterns and employment biographies. Improving social benefits for part-time workers, decoupling rights from marital status, and granting social benefits for child care duties are typical examples of recalibrating compensation (for examples in the area of pension policy in various European countries, see Bonoli, 2000; Immergut et al., 2007; Häusermann, 2010; Palier, 2010). Recalibration benefits mostly women and more generally people with discontinuous employment biographies. It is presumed to find support among those groups and among advocates of universalistic, gender-egalitarian social policies more generally.

Finally, the third strategy of compensation involves combining benefit cutbacks with *increased revenues* for social security in order to bolster the financial stability of social security in the longer run and to share the burden between both sides of expenditures and revenues (for examples of contribution increases as a reform strategy in many countries in continental Europe, see Palier, 2010). Even though contribution increases (particularly through pay-roll taxes or VAT) also represent a cost to voters, we still consider them a compensation strategy, since balancing financial consolidation between taxes and expenditures has been a long-standing claim of the left in many countries. Hence, this compensation strategy should bolster support for the reform particularly among left-wing voters.

In this paper, we test the effect of including such compensations on overall support for

a reform package that contains unpopular retrenchment. Before developing more specific expectations, we state a first general hypothesis (H1) that compensation does increase popular support for welfare reform. More specifically (in terms of an observable implication), we expect to find that reform packages, even though they may contain elements that are strongly contested on their own, will nevertheless receive majority support when compensating measures are included. One may ask if this mechanism applies only to citizens who agree with the premise that cutbacks are necessary. Citizens who do not believe in the necessity of hard choices and trade-offs may reject any retrenching reform, irrespective if it contains compensations or not. If that was indeed the case, our argument would only hold within the scope condition of this premise.[5] Our survey allows us to empirically address this question below. However, on a theoretical basis, there are grounds to think that elements of compensation may be effective in fostering support among all citizens: targeting, recalibration and increased revenues are not simply limitations on cutbacks, but they are policy reform elements on their own. In other words: if compensation was simply a limitation of cutbacks for some categories, we would not expect them to enhance support among those who deem cutbacks non-necessary in the first place. However, since compensation elements affect support on distinct dimensions, they have their own value and meaning, (to some extent) independently from the cutbacks.

Related to this, the more specific question, of course, concerns *which compensation strategy is effective among which voters*. To motivate our hypotheses, we rely on recent research on the relevance of material approaches for the explanation of social policy preferences. While the overall evidence on this link is mixed (Mutz, 1998), recent studies have found strong support for a link between the probability of an individual to become a recipient of a social benefit and her propensity to support this benefit (Rehm, 2016), especially in the context of mature welfare states (Busemeyer & Neimanns, 2017). In other words, while ideological or predispositional determinants matter most for attitudes regarding the overall level of social spending and distributive principles, material

[5]In Switzerland, this would only be a narrow restriction to our argument since only 12.1% of our respondents agree with the statement that "Reform pressure is exaggerated. Cutbacks are not necessary", while 88% deem cutbacks "necessary" or "urgently necessary".

interest is indeed relevant when it comes to explaining preferences regarding the specific distribution and design of benefits. Given that targeting, recalibration, and increasing tax revenues are directed toward lower-income voters, women, and left-wing voters respectively, we will thus study the impact of income, gender, and partisanship on reform support. We expect people with lower incomes to be particularly responsive to targeting (H2) and women to be particularly responsive to recalibration (H3). In other words, we expect targeting to contribute to the support for a reform package among lower-income people, and we expect recalibration to increase the support for a reform among women. In addition, we expect voters of left parties to be particularly responsive to recalibration, targeting, and increased revenues (H4) since all of these compensation strategies address key concerns of left-wing political programs (i.e., the correction of inequalities, societal modernization, and equal burden-sharing).

3 Case Selection, Experimental Design, Data, and Estimation Methods

3.1 Pension retrenchment in a direct democratic context

We test the effects of different compensation strategies in the field of pension policy reform in Switzerland. This case provides us with an ideal real-world setting to study welfare reform attitudes.

Pension reform is the prime example of a welfare policy in which reform is constrained by public opinion because mature pension regimes in the developed OECD countries feature precisely the endogenous institutional feedback effects that account for widespread opposition against retrenchment. Hence, pension policy is most likely the area in which political exchange and compensation are key conditions for reform success (Immergut et al., 2007). Pension policy is also an ideal policy area for comparing the various compensation strategies because old-age income protection relies on a range of different distributive principles (Pierson, 2001b; Palier, 2010). Basic pension schemes provide oppor-

tunities for targeting reforms. Earnings-related occupational pensions generally lead to pension differentials between men and women. Moreover, this gender bias has entailed new social risks for women that create a need for recalibration. Finally, most countries finance public pensions through contributions or taxes, which implies that governments can indeed rebalance pension finances either via reduced benefits or via increased revenues.

We acknowledge that old age pension policy is a field for which our argument regarding the effectiveness of compensations may apply more aptly than to other fields. Since substantial cuts are difficult to nuance and therefore affect a wide share of the public, they tend to be highly visible and politically salient. At the same time, old age pension systems provide ample grounds for compensation of particular groups. If cuts could indeed be obscured and hidden from the public, compensation would of course be less relevant. One may think that expenditure cuts in services (e.g. health) may be more easily obscured. In that sense, the visibility of cuts is a scope condition for our argument, as it presupposes that actors politicize cutbacks.

Switzerland provides an ideal empirical setting for studying policy reform attitudes because survey respondents are confronted with a realistic task. Since Switzerland is a semi-direct democracy, major pension reforms are usually subject to a direct democratic referendum.[6] Therefore, most respondents have evaluated policy trade-offs and compensations in the past. Indeed, over the past 25 years, Swiss voters have been called to the polls no less than seven times to vote on pension reform (Häusermann, 2010). These context conditions improve the external validity of our experimental findings because the survey puts respondents in a real-world situation.

In order to further improve the validity of our findings, we study reform attitudes with regard to an actual policy reform process rather than asking respondents about abstract or hypothetical preferences, as is usually done in similar surveys. Measuring preferences without putting them in a relevant real-life context can create distortion for various framing or confounding effects (Bertrand & Mullainathan, 2001). The most recent

[6]Referenda on reforms of laws are optional in Switzerland. If opponents of the reform manage to collect 50,000 signatures in 100 days, the reform will be voted on. Despite not being mandatory, major social policy reforms are almost always subject to a referendum.

Swiss pension reform ("Altersvorsorge 2020") provides the perfect example of a policy reform that was aimed primarily at financial consolidation but that was supposed to foster popular support by compensation.[7] The government's explicit aim in this reform was to secure the financial stability of the pension regime until at least 2030. To this effect, the government proposed in 2014 to raise the retirement age for women and to lower general pensions as well as widows' pensions. Combined, these two measures would reduce expenditures massively by almost six percent. The estimated savings by 2030 compared to the current annual expenditure of the system are presented in Appendix B in the supplementary information. In order to mitigate the effects of such retrenchment and to foster popular support for the reform, the government also proposed several expansive elements to be included in the same reform. Table 1 provides an overview of the most important, visible, and politicized reform elements in the government proposal and theorizes them in the terms developed above. It is important to notice, however, that despite the compensating elements, the overall effect of the reform would still have been to reduce spending. Even if the proposed retrenchment were compensated at the maximum rate, the proposed reform would still have curbed expenditures by at least 2.5 percent.

[7]The Swiss pension system provides opportunities for all three strategies of compensation: the first pillar (AHV) provides universal, redistributive basic public pensions. By contrast, the second pillar provides occupational pensions on a strictly actuarial basis (social insurance). Hence, while the first pillar provides ample opportunities for targeting compensations, the Achilles heel of the second pillar typically is a bias against female employment biographies, which brings a demand for recalibration to the agenda. Since part of the first pillar is financed via VAT, this provides leeway for increasing revenues.

Table 1: The Swiss pension reform "Altersvorsorge 2020"

Reform elements	Content	Goal of the reform elements
Pension cutbacks 2nd pillar	The conversion rate denotes the rate at which the capitalized old-age savings in the second pillar are calculated into annual pension benefits. Currently, this rate is 6.8%. The bill proposed a reduction to 6%.	**Retrenchment**
Cutbacks in widows' pensions	Switzerland has generous widows' pensions for all married women below the age of 64. The reform proposal suggested that in the future only widows with dependent children should receive such a pension.	**Retrenchment**
Increase in age of retirement	Currently, the age of retirement in Switzerland is 65 for men and 64 for women. The reform proposed to equalize the retirement age for both men and women at 65. Right-wing parties demanded a further increase to 67.	**Retrenchment**
Subsidies for early retirement	Currently, early retirement implies a significantly lower pension level (linear cutbacks proportional to the years of early retirement). The bill proposed to subsidize early retirement for low-income workers only.	**Compensation: targeting**
Extended eligibility 2nd pillar	The second pillar of occupational pensions in Switzerland insures only incomes above ca. 24,000 CHF/year. This is detrimental to workers who work part-time or combine several part-time contracts. As part-time work is very predominantly female in Switzerland, this issue affects mainly women. The reform bill proposed to lower the access threshold, so that about 150,000 women would be newly insured.	**Compensation: recalibration**
Increased revenues (VAT)	Pensions in Switzerland are financed by about 75% via payroll contributions. In addition, the government contributes to financing with revenues from income taxation and from VAT. The government suggested increasing the amount of VAT earmarked for the basic pension system.	**Compensation: increased revenues**

3.2 The experimental design

Studying the "politics of trade-offs" required that we identify the relative importance of different reform elements. When a reform package contains both, elements that a voter rejects and, elements that he/she favors, the voter will perform an (implicit) balancing of the relative preference importance, which eventually determines whether he/she supports or rejects the reform package as a whole. Standard surveys do not allow us to measure relative preference importance (for a similar critique, see Busemeyer & Garritzmann, 2017). Generally, support levels are very high for expansive social policies and very low for retrenchment. However, from this information, we cannot evaluate public opinion about actual welfare state reforms since these reforms are usually multidimensional. In addition, standard survey questions (even if they ask about priorities) artificially separate

attitudes toward individual elements of policy change. However, in actual preferences concerning reforms, these elements are linked, and a voter may evaluate a measure differently depending on the other elements of the reform. Therefore, a standard random utility model that infers overall utility from several separately measured components is unable to estimate overall utility correctly (Horiuchi, Smith & Yamamoto, 2015).

We used conjoint survey analysis as a more appropriate tool to analyze current welfare politics because it precisely prompts respondents to choose between different policy packages. It is therefore perfectly suited to examine individual preferences in the context of multidimensional reforms. Moreover, the conjoint design has been shown to approximate real-world decisions more closely than vignette designs (Hainmueller, Hangartner & Yamamoto, 2015). Conjoint designs have a long history in psychology, marketing, and sociology (Green, Krieger & Wind, 2001; Wallander, 2009), but they have only recently started to be used in political science, with a few pioneering studies showing the substantial value for political science research questions (Bechtel, Hainmueller & Margalit, 2014; Hainmueller & Hopkins, 2014; Bechtel, Genovese & Scheve, 2017; Gallego & Marx, 2017; Ballard-Rosa, Martin & Scheve, 2017) and providing methodological support for applications in political science (Hainmueller, Hopkins & Yamamoto, 2014; Hainmueller et al., 2015).

We designed and implemented our experiment as follows: First, we specified the core elements of the reform package (six in total, see Table 1) and different values for each reform element (levels). To give an example, the element "Increase in age of retirement" can take three levels: 1. Status quo: no increase; 2. Increase by 1 year for women; 3. Increase to 67 years for both, men and women (see Table 2 for all possible levels of each reform element).

The definition of reform elements and levels is, obviously, key for both the internal and external validity of the experiment. In our case, we identified six core elements of the ongoing reform and we defined three levels for each element, according to the same principle: status quo, government proposal, claims to go beyond the government proposal. Table 2 shows the design. It was important for us to include only levels that are realistic

and that were actually discussed in the public debate, which is why we defined only two levels for recalibration.[8] One downside of the realistic setting for our survey could be that respondents compare the hypothetical packages not only between themselves, but also to the "real" one. However, our survey was fielded before the parliament had discussed and decided on the reform, i.e. before a fully defined reform proposal was present in the public debate. Also, the robustness survey that we fielded one year later (see section "Robustness Checks" below) – during a time when the debate about the full, real reform proposal was ongoing – showed no signs of this change affecting the findings.

In a next step, reform packages were generated randomly. They contained the whole set of reform elements (in random order) and a random selection of levels for each reform element. Survey respondents were then presented with two hypothetical reform packages to compare and evaluate. More specifically, they were asked to indicate (a) which of the two reform packages they preferred (the "choice" variable) and (b) how likely they would be to support each of the two reform packages individually in a popular referendum (the "ranking variable", similar to Bechtel et al., 2014) (see Appendix A for a screenshot of the conjoint survey). Through randomization and a high number of such pairwise comparisons, this empirical strategy allows us to identify (and quantify) the causal effect of including specific levels of individual reform elements (e.g. increase in age of retirement for women by 1 year) on the support for the entire reform package, compared to the support for a reform package that contains the baseline level (status quo) of this particular reform element. In other words, this allows us to estimate the effect a deviation in one reform element from the status quo has *ceteris paribus* on the support for the whole reform package.

Each respondent compared five pairs of hypothetical reform packages[9], and had to make a decision between the two of them before being able to continue the survey. All

[8] In our experiment various combinations of levels are possible. The design could exclude certain combinations that are impossible or unlikely (at the cost of the benefits of randomization); however we decided not to exclude any combinations, because all packages are politically possible. The only restriction we imposed prevented the display of two exactly identical packages.

[9] To prepare respondents for their task, the survey questionnaire started with multiple screens that explained the overall structure of the pension system and the context of the reform as well as the meaning of the reform elements.

Table 2: Reform elements that are being discussed, levels and goals

Reform elements	Levels		Goal of the reform elements
Pension cutbacks 2nd pillar	1: status quo	No cuts (6.8% conversion rate)	**Retrenchment**
	2: government proposal	Cutbacks; balance the lowering of pension levels by having people contribute more	
	3: beyond gvt	Cutbacks; no balancing	
Cutbacks in widows' pensions	1: status quo	All widows under 64 eligible for benefits	**Retrenchment**
	2: government proposal	Only widows with children <16 years old would be eligible	
	3: beyond gvt	Stepwise abolishment of widows' pensions	
Increase in age of retirement	1: status quo	64 for women, 65 for men	**Retrenchment**
	2: government proposal	Increase for women by 1 year: 65 for both	
	3: beyond gvt	Stepwise increase for both men & women to 67	
Subsidies for early retirement	1: status quo	Early retirement allowed, but with linear cutback in the benefit level	**Compensation: targeting**
	2: government proposal	Early retirement subsidized for lower-income earners	
	3: beyond gvt	Early retirement subsidized for all	
Extended elibility 2nd pillar	1: status quo	No change; only people earning >24,000 CHF/year are eligible	**Compensation: recalibration**
	2: government proposal	Extend access for people with lower incomes and part-time workers	
Increased revenues (VAT)	1: status quo	No increase in VAT	**Compensation: increased revenues**
	2: government proposal	Increase VAT by max. 1.5 percentage points	
	3: beyond gvt	Increase VAT by max. 3 percentage points	

respondents were also forced to indicate their level of support for each reform package in a hypothetical referendum. Our design resulted in more than 18,000 choices and ratings (1,873 respondents x 5 comparison x 2 reform packages) of randomly generated reform packages.

Given the complexity of the task, two concerns regarding internal and external validity need to be addressed: the specificity of the elements and levels and respondent fatigue. Regarding the specificity of the levels, one may first ask if the reform elements we included were so specific that they limited the external validity of the findings because it would be unrealistic to assume that citizens would know about specific reform elements in other countries. We suggest that our findings *do* apply beyond the context of Switzerland because public opinion and debate about welfare reform is *always* specific and focused

on particular measures (e.g., Articolo 18 in Italian labor law or fixed-term contracts (CDD) in France – two highly specific elements that became decisive elements of political contestation). Beyond this general observation, we have taken several measures to ensure that we include only those reform elements that were actually politicized and visible. We defined reform elements and levels based on the government report accompanying the reform proposal as well as on empirical analyses of the key elements in previous pension reforms and pension votes. We then discussed our design with public servants from the Federal Office of Social Insurances (i.e., the competent administrative office). Furthermore, we implemented both a qualitative pretest of respondents as well as a quantitative pretest of 150 students. Finally, we verified that all reform elements and their levels (including even the more technical ones) were discussed prominently in the media, based on data from the FOEG media observatory.[10] Based on this procedure, we feel confident that despite the task being specific, it adequately reflected opinion formation on this reform in the public.

The second concern regarding the complexity of the reform elements and levels is whether only highly sophisticated voters/respondents would understand the reform proposals and be able to interpret them as retrenchment/compensation or whether less sophisticated voters would answer randomly or react to obfuscation rather than to compensation. This is a concern that we addressed empirically first, by cross-validating our conjoint findings with simpler direct questions; second, by controlling for education and empirically comparing the consistency in the answers of high- and low-education respondents; and third, by comparing the coefficient sizes of the more technical and the more straightforward elements (in order to test for obfuscation). The findings of these robustness tests (Appendix F) confirm that the levels and the survey task was understood consistently. We discuss the findings of the robustness test in more detail in Section 5.

Regarding respondent fatigue after the first tasks, we contained this risk in three ways: First, as already mentioned, we performed extensive quantitative pretests (with students) as well as qualitative pretests in which respondents were observed via camera while they

[10]FOEG is the Forschungsinstitut Öffentlichkeit und Gesellschaft at the University of Zurich.

filled in the questionnaire (the respondents had been asked to articulate their thought-process while responding). This pretest helped us design the most helpful instructions for all five paired comparisons. Second, we measured time stamps throughout the survey, in order to detect and exclude respondents who clicked through the comparisons too quickly (less than 3 percent of the respondents). Finally, we estimated all results based only on the first two (out of five) conjoint comparisons, assuming that people's concentration was strongest at the beginning of the task. The findings are entirely robust (see Appendix C).

3.3 Data and estimation strategy

The survey was conducted by the social research company gfs.bern in both the French- and the German-speaking parts of Switzerland among a representative sample of Swiss citizens between March and June 2015 (after a pretest in February).[11] The sample includes 1,873 fully completed interviews. Random sampling was done on the basis of the national telephone register, which indicates the first phone number an individual registered (mobile or landline number). Respondents were recruited via CATI interviews in which they indicated their e-mail address. They then received an e-mail within 24 hours with an access code for the online survey. Respondents were reminded three times as needed (via e-mail after two and after three weeks, and via telephone after four weeks). Our sampling strategy was based on quota for the region, age, and gender drawn from the national census. We chose to oversample respondents over 65 because this age group usually has lower participation rates in online surveys. Table 3 shows sampling and response rates by region, age, and sex. Our overall response rate was 63 percent; it was highest among elderly men and lowest among young men and in the French-speaking part of Switzerland.

In the analyses, we account for remaining bias from survey response using weights created by iterative post-stratification. The weights adjust for region, age, and gender

[11]The exclusion of the Italian-speaking canton of Ticino from this survey was based on purely pragmatic, cost-related consideration. Its inclusion in the second robustness-survey one year later (see section 5 below) showed no impact on the findings.

Table 3: Response rates by language region, age, and sex

Region	**Recruited**			**Login**			**Completed**		
German	2307			1720			1570 (68%)		
French	637			340			303 (45%)		
Total	2980			2060			1873 (63%)		
Age	**Recruited**			**Login**			**Completed**		
	m	**f**	**all**	**m**	**f**	**all**	**m**	**f**	**all**
<40	390	379	769	168	247	415	157 (43%)	209 (55.1%)	366
40-64	572	611	1183	423	447	870	392 (68.5%)	399 (65.3%)	791
>65	462	566	1028	361	414	775	341 (73.8%)	375 (66.3%)	716
Total	1424	1556	2980	952	1108	2060	890 (63.5%)	983 (63.2%)	1873

(based on known population margins) and for party preference (based on the national election results). Following the empirical strategy of Hainmuller et al. (2014), we estimated the average marginal component effect (AMCE), which is the marginal effect of a specific level of a reform element averaged over all possible combination of levels of the other reform elements given their joint distribution.[12] [13] Because each level was randomly assigned, we were able to estimate the AMCEs using a simple linear regression of the following form, where every respondent (indexed as i) is presented with K conjoint comparisons (k) and in each comparison chooses one of J alternatives (j):

$$\begin{aligned}\text{chosen}_{ijk} = \theta_0 &+ \theta_1[\text{element1}_{\text{ijk}} = \text{level2}] + \theta_2[\text{element1}_{\text{ijk}} = \text{level3}] \\ &+ \theta_3[\text{element2}_{\text{ijk}} = \text{level2}] + \theta_4[\text{element2}_{\text{ijk}} = \text{level3}] \\ &+ \theta_5[\text{element3}_{\text{ijk}} = \text{level2}] + \theta_6[\text{element3}_{\text{ijk}} = \text{level3}] \\ &+ ... + \epsilon_{ijk},\end{aligned}$$

where $\text{chosen}_{\text{ijk}}$ is coded as 1 if that hypothetical reform package was chosen and

[12] We used a uniform distribution, see FN 8 above.

[13] Note the difference between conjoint analysis and the average treatment effect (ATE) – the expected difference in response between treated and untreated – in standard survey experiments: Common survey experiments identify the causal effects of the treatment *as a whole*, that is, we cannot know which element of the treatment produced the effect, if the experimental manipulation included several attributes (Hainmueller et al., 2014). Coinjoint experiments are designed to study multidimensional decision-making: they allow us to identify the causal effects of the different components of a treatment. The estimated parameter, the AMCE, differs from the ATE in common survey experiments insofar as it measures the marginal effect of *each treatment component*, in our case the levels – or the changes from the status quo – independently of the configuration of the other components (Hainmueller et al., 2014).

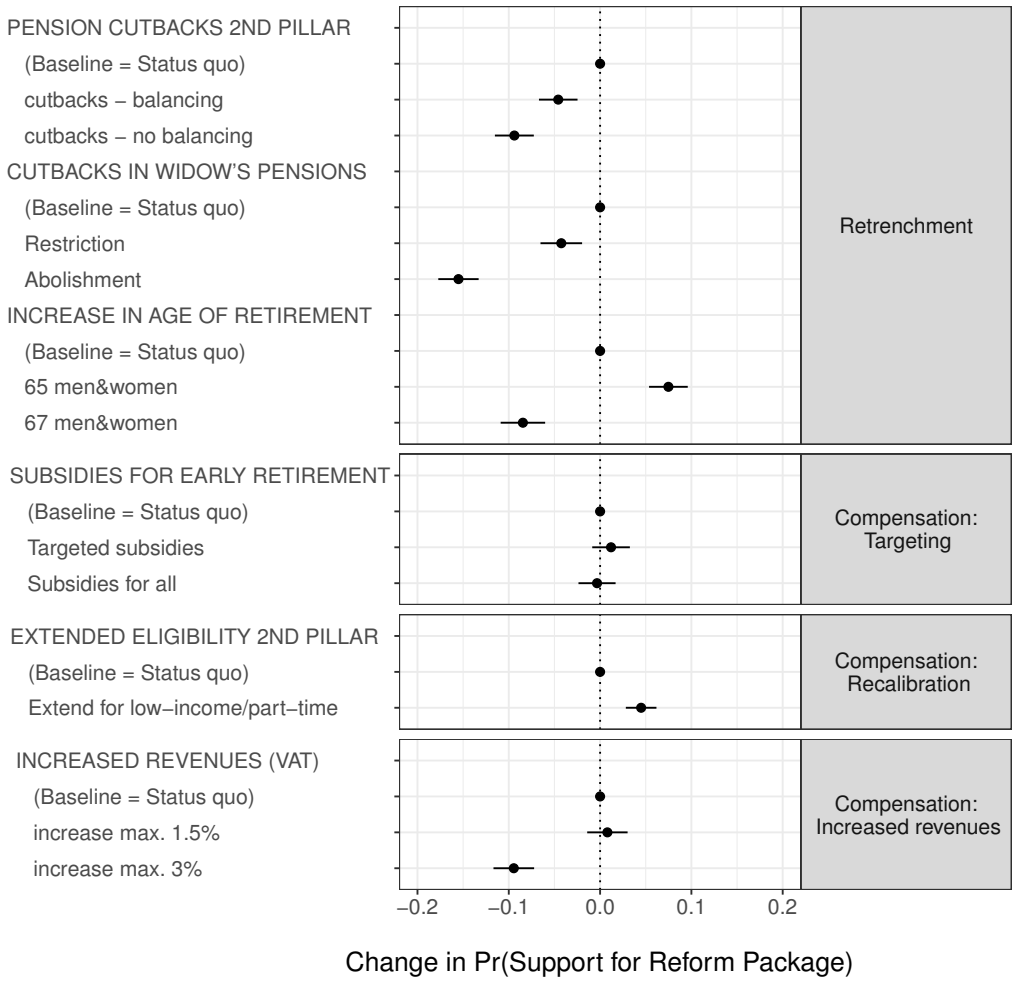
0 otherwise. [element3$_{ijk}$ = level2] and [element3$_{ijk}$ = level3], for example, are dummy variables coded 1 if the reform element 3 (Increase in age of retirement) took the respective level in the shown policy package and 0 otherwise. Level 1, the reference category, is always the status quo. Accordingly, $\hat{\theta}_5$ estimates the average difference in the probability of a package being chosen if it raises women's retirement age by one year as compared to the status quo (retirement age = 65/65), with the average being computed over all possible combinations of levels of the other reform elements. The analysis of the conjoint data was done with the `cjoint` package in the `R` language (Hainmueller et al., 2014). Clustered standard errors account for multiple pairwise ratings of reform packages by each respondent.

4 Findings: Effects of Retrenchment and Compensation on Support

Our first hypothesis states that compensation is effective in counterbalancing opposition against retrenchment. To gauge this effect empirically, we examined the findings of the conjoint analysis for the full sample of respondents. Figure 1 first lists the three retrenchment elements of the reform package (cutbacks in benefit levels of second pillar pensions, cutbacks in widows' pensions, and a higher retirement age) followed by the different compensating elements. The coefficients indicate for each level of a reform element the average change in the probability of supporting a reform package if the package includes the respective level instead of the baseline level (status quo). Negative coefficients indicate that a specific level would reduce overall support.

As expected, we see that pension retrenchment (pension cutbacks, lower widows' pensions, and a general increase in the age of retirement) clearly, significantly, and sometimes massively reduces support for the reform. Raising women's retirement age from 64 to 65, by contrast, increases support for the overall reform. It seems that most voters do not perceive this change as retrenchment, but rather as a measure of updating/gender

Figure 1: Effects of reform elements on support for the pension reform package



Note: Estimates shown are changes in probability of supporting a pension reform that includes the specific level of a reform element, compared to a reform that includes the status quo. Bars represent 95 percent confidence intervals.

equalization. Hence, a first finding confirms that—even when presented in a reform package—retrenchment elements considerably lower the chances of the reform. Can expansive elements compensate for this loss of support? When we compare the average effects of different compensation strategies, we see that their effects differ strongly: Subsidies for the lower-income earners in favor of their possibility to retire early are a form of targeting compensation. In our sample, however, they do not increase the support for the entire reform significantly. Even though the coefficient is slightly positive, this reform element is not salient enough in the entire sample of respondents to reach significance. The same holds for the compensation via increased revenues from VAT (1.5 percent): a more massive tax increase (3 percent) reduces the chances of the reform package even more

clearly. The compensation strategy that significantly increases support for the entire package in the full sample of respondents is recalibration: extending eligibility to second pillar pensions for low-income earners and part-time workers (mostly women) increases the probability of a reform being accepted by almost 5 percentage points. In other words, this compensation is able to counterbalance the cost of cutbacks (e.g., the lowering of second pillar pensions).[14]

Alternatively, we can look at the effectiveness of compensation from a different perspective. Due to randomization of the levels, our design involves both reform packages that include only a small amount of retrenchment and some very painful packages that consist entirely of cutbacks. The underlying assumption of our design is that the overall support for a reform package decreases the more retrenching elements it includes. Put differently, as retrenchment implies lower expenditures, we would expect the popularity of a reform package to be negatively related to the savings potential of a reform. The real-world setting of our design allows for testing the validity of this assumption. We computed the aggregate cost implications of all 486 hypothetical reform packages on the basis of the budgetary implications of the reform elements included (see Appendix B). Cost implications range from the most restrictive packages (only retrenchment, no compensation), which would lead to a 10.6 percent expenditure reduction (9.8 bn/year), to the most expansive ones (no cutbacks, only compensations), which would expand expenditures by 0.8 percent per year (0.8bn/year). On the basis of the ranking variable ("how likely would you support this package in a referendum?"), the histogram in Figure 2 presents the distribution of acceptance rates (dichotomized ranking variable) for each of the 486 packages. As expected, the acceptance rate varies widely, with some unpopular combinations of reform elements only supported by 20 percent of respondents or less and some packages being supported by an overwhelming majority. The four foregrounded distributions (kernel densities) in Figure 2 take into account the available information on

[14]One may argue that the idea of compensation presupposes that expansive elements are presented in combination with retrenchment. Since all level combinations are possible, we also have packages that are purely expansive in our sample. To test for robustness, we calculated the findings based on a sub-sample of hypothetical packages that includes only those reforms that contain any kind of retrenchment. The findings are robust (see Appendix C).

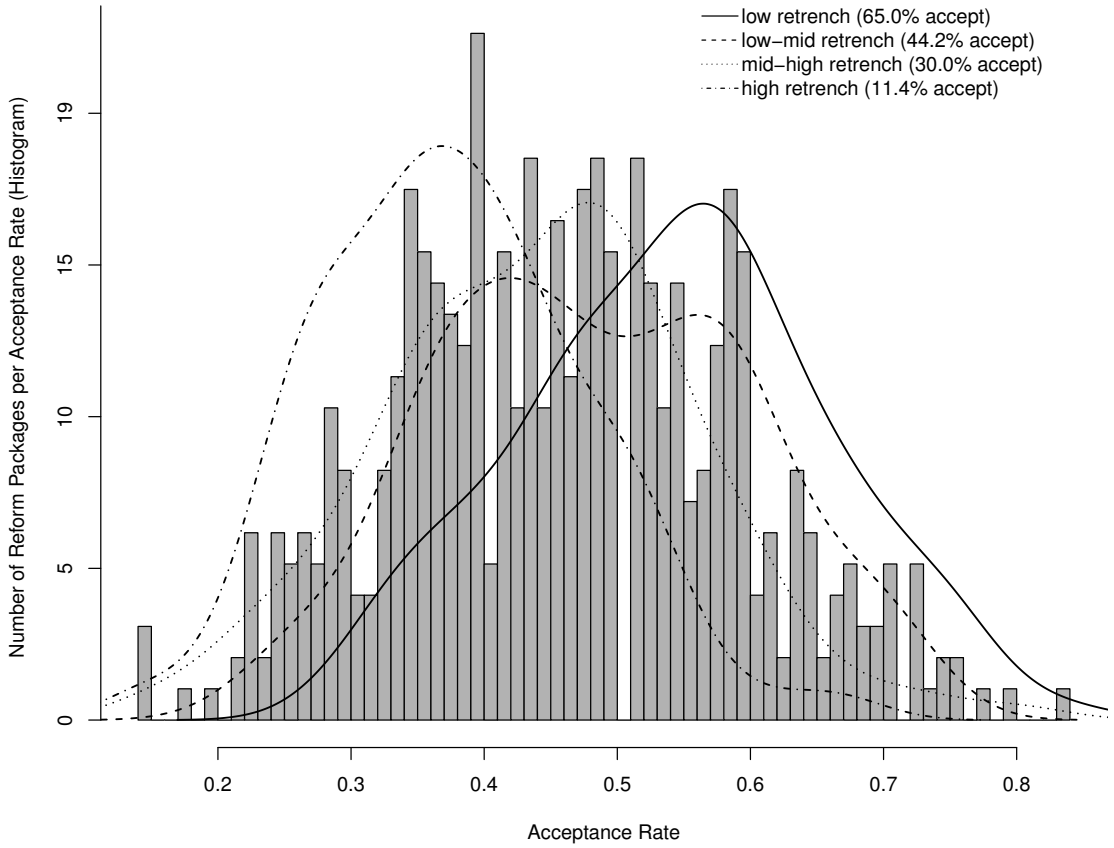
the savings potential of each reform element and demonstrate that more restrictive packages receive less popular support. To generate these distributions, we ranked the reform packages according to their saving potential (i.e., the severity of retrenchment) and split them into quartiles. The message is clear: the more retrenchment a package involves, the lower the share of respondents supporting it. However, an observable implication of our hypothesis 1 is that we should also find retrenching reform packages (i.e., packages that imply a substantial decrease of government expenditures) that would receive a majority of votes *despite* the included cutbacks. This is indeed the case. Even among the second and third quartiles (retrenching pension expenditures by 2.3 to 6.3 percent annually), we find that a substantial share of these reforms receive acceptance rates above 0.5, which means they would enjoy popular support of at least 50 percent. Reform packages that include very weak or no compensation (the fourth quartile), however, hardly ever gather majority support and thus would not stand a chance in a popular referendum. In terms of scope conditions, it is important to note that this finding holds for all respondents, irrespective of whether they agree with the premise that reform pressure is high, or that cutbacks are necessary.[15]

Hypothesis 2 suggests that lower-income people should respond particularly strongly to compensations that *target* their specific pension rights. Subsidies for early retirement are such a compensation, but the extension of the eligibility to second pillar pensions also has some (limited) traction for lower-income earners' pensions. Accordingly, Figure 3 shows the conjoint findings for individuals living in lower-income households[16] compared

[15]Our survey includes an item measuring perceived reform pressure. Respondents were asked "How do you perceive the reform pressure in Swiss old age pension system? How urgent are restrictive reforms?" and they could agree to one of the following statements: "Restrictive reforms are necessary and urgent. Demographic ageing is an immediate threat to the stability of Swiss old age security.", "Restrictive reforms are necessary, but the situation is not dramatic. The challenge of demographic ageing can be handled financially via appropriate reforms." or "Reform pressure is exaggerated. Cutbacks are not necessary". Only 12% agreed with the last statement, whereas about 55% agreed with the middle category and about a third of the respondents agreed that reform pressure was high and urgent. This item allowed us to test whether our argument on the effectiveness of compensation holds only among those who accept the premise of reform pressure, whereas the others might be immune to any compensation. The pattern shown in Figure 2 holds for all categories of respondents, even those who perceive low reform pressure. This provides evidence that our findings hold on a very broad basis.

[16]We define lower-income individuals as those living in households with a combined income of below 6,000 CHF/month. This corresponds to the median individual income in Switzerland and represents 1.5

Figure 2: Support for reform packages depending on cost implications



Note: X axis: share of positive responses to the question: "If you had to vote on package X in a popular referendum, would you support the reform or reject it?" Left Y axis denotes the bars, right Y axis the lines. Lines: all hypothetical reform package combinations (total of 486) have been divided into quartiles, depending on their impact on the reduction of expenditures (retrenchment). The cutoff points of the four quartiles are expenditure cuts of 5.79 bn/year, 4.18 bn/year, and 2.07 bn/year. Cost implications are based on the official estimates as reported in Appendix B.

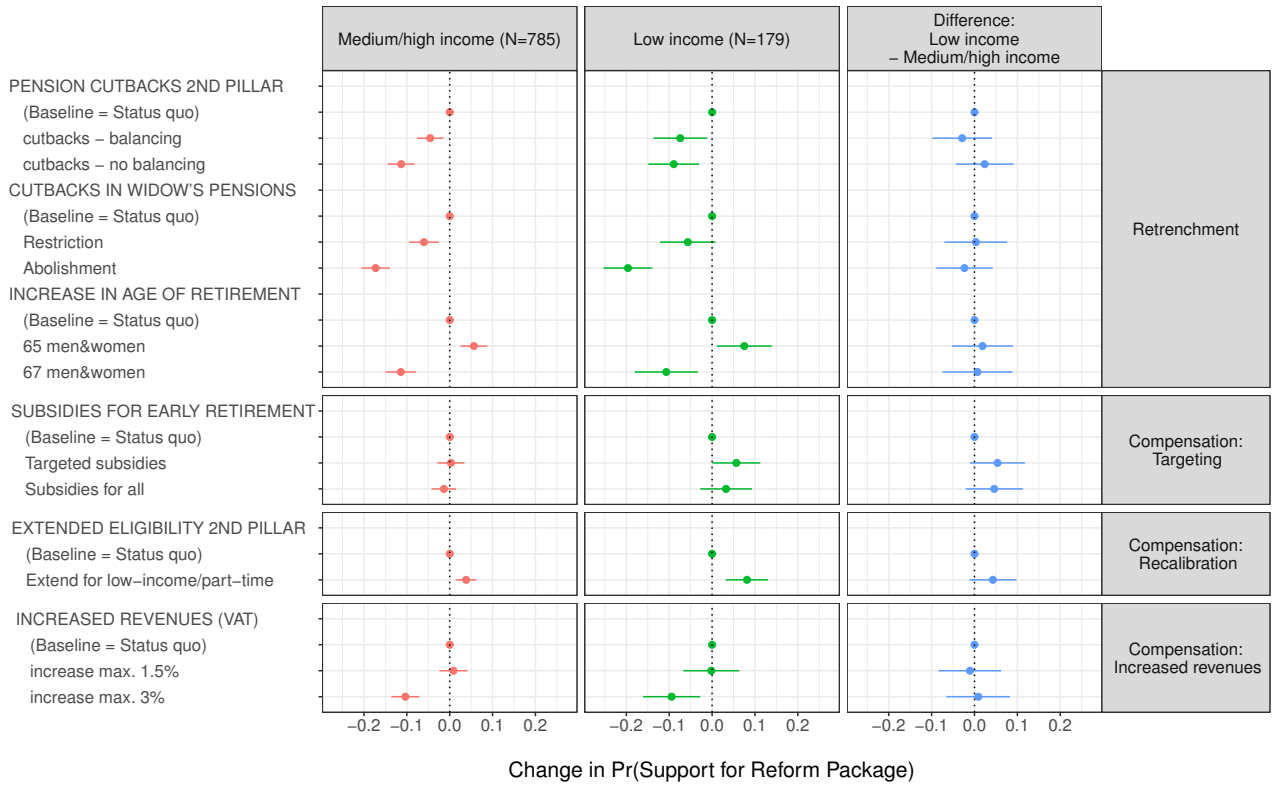
to higher-income classes.

We found that among lower-income earners, targeted subsidies indeed increased support for the reform by 5.7 percentage points, while the middle- and higher-income earners were basically indifferent towards these reform elements. However, even though the effect was significant among the lower-income earners, the difference between the income groups was not (see right panel of Figure 3; Appendix D provides the full list of coefficients).

A second form of compensation, *recalibration*, more strongly increases support for the

times the threshold the government defined for individuals to be eligible for subsidies for early retirement. Pensioners are excluded since their pensions would not be affected by the compensations.

Figure 3: Effects of reform elements on support for the pension reform package by income



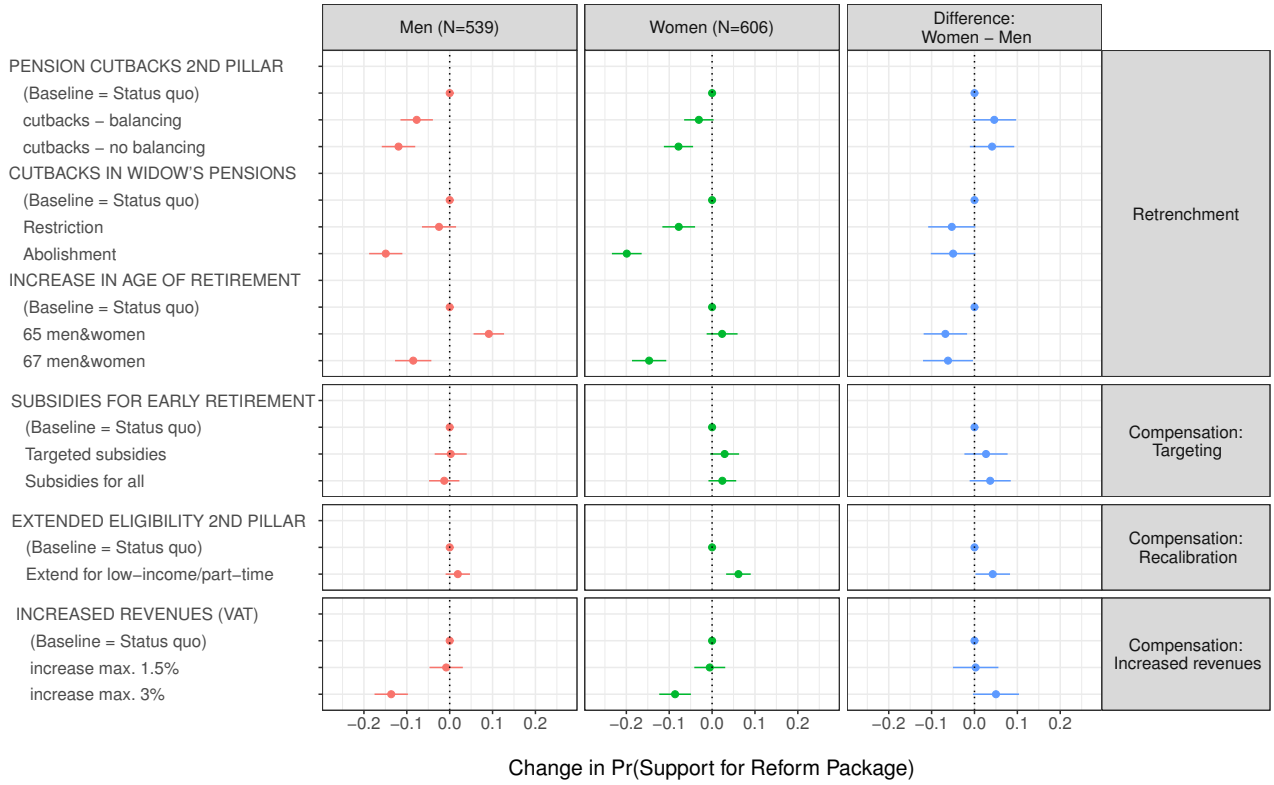
Note: Estimates shown are changes in probability of supporting a pension reform that includes the specific level of a reform element, compared to a reform that includes the status quo. Bars represent 95 percent confidence intervals.

reform packages among lower-income earners compared to higher-income groups. Again, the difference between the two subgroups is not significant. Hence, income as such does not seem to be a very strong predictor of the specific relative importance different compensations have for different groups. Attitudes among lower-income earners are not sufficiently distinct from those of higher-income groups to make such a claim.[17]

Differences are stronger when we compare active men and women (pensioners are excluded here since the cutbacks would not affect them) based on their relative material interests in the compensations offered. We suggested that if material interests drive the effectiveness of compensations, women should be more responsive to recalibration than men since the extension of eligibility to second pillar occupational pensions affects

[17]Nevertheless, of course, compensations do work among this group: 27.5 percent of reform packages that include pension cutbacks in the second pillar still receive majority support among lower-income individuals who explicitly reject such cutbacks if they include recalibration measures. However, this pattern is not distinctive enough to support income as a determinant of compensation effectiveness.

Figure 4: Effects of reform elements on support for the pension reform package by gender



Note: Estimates shown are changes in probability of supporting a pension reform that includes the specific level of a reform element, compared to a reform that includes the status quo. Bars represent 95 percent confidence intervals.

mostly women. About 85 percent of employed women in Switzerland work only part-time. Extending access to the second pillar as proposed by the government would have newly granted occupational pension rights to more than 150,000 women. Indeed, recalibration (extending eligibility for second pillar pensions) increased support for the reform significantly by 6.1 percentage points among women, while the effect was small and not significant for men (Figure 4). Unlike the difference between income groups, the difference between men and women with regard to recalibration was significant (see Appendix D in the supplementary information). Overall, it appears that compensating women is particularly important for the success of reform since some of the gender-specific restrictive elements of the reform (e.g., lowering widows' pensions, increasing women's retirement age by one year) generate stronger resistance against the entire package among women than among men. Regarding the government proposal (see Table 1), men oppose only pension cutbacks, which would affect them directly, while they are supportive of increas-

ing women's retirement age, and they are indifferent with regard to restricting widows' pensions. Hence, it seems that among men, such a reform overall should generally have a better chance of being supported than among women. Indeed, the government seemed aware that women's votes would have to be won in this process because it explicitly argued that since several of the cutbacks have greater impact on women than on men, women's overall pension rights should be particularly improved in the second pillar.[18] According to our results, this strategy is indeed effective: overall, among those women who reject pension cutbacks explicitly, 25.5 percent of the reform packages that contain precisely such cutbacks would nevertheless receive majority support.

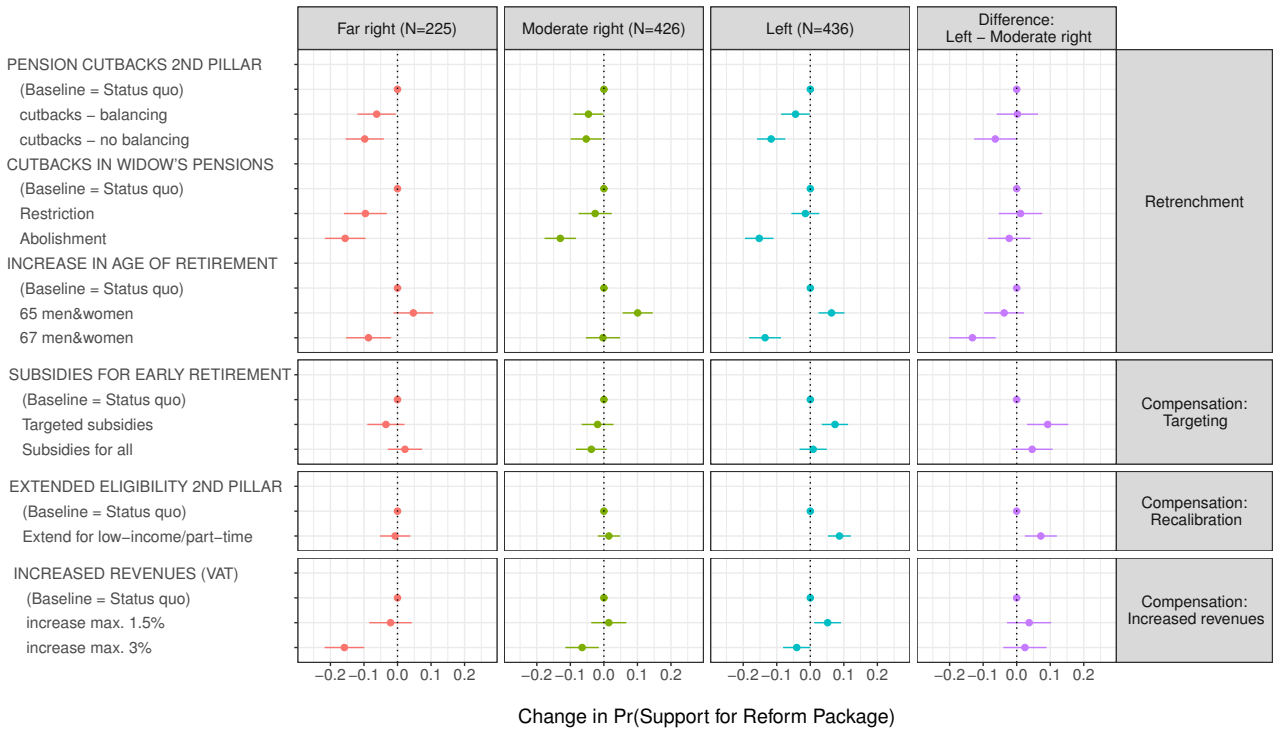
Finally, we tested partisanship as a determinant of the effectiveness of different compensation strategies. In all past referenda, it was the Social Democrats and/or trade unions that launched the referenda against reforms and led the campaigns (of course, to win the referenda required far more votes than just those of the left-wing voters). Hence, for a government seeking to strengthen support for a reform, it seems particularly important to foster support among those on the left.

We suggested that retrenchment should reduce support more strongly among left-wing voters than among center-right voters; in addition, H4 posits that recalibration, targeting, and increased revenue should increase support for the reform among left voters more strongly than among center-right and far-right because they address long-standing claims that the Left has raised regarding the expansion of old-age income security.

Figure 5 compares attitudes of left-wing voters (Green Party and Social Democrats) to those among voters of the moderate right (Christian Democrats and Liberals) and voters of the far right (Swiss People's Party). The data in the right-hand panel shows how significantly left-wing voters differ from moderate right voters. These parties are the main opponents in this policy field. The Far right party elites support retrenchment, but their voters do not, which is why the SVP tends to avoid the spotlight in the reform process. First, we found that pension retrenchment reduced the overall support for the reform among all voters (except for the equalization of the retirement age for men and women).

[18]Faktenblatt: Altersvorsorge 2020. Was die Reform fur die Frauen bedeutet. Bundesamt fuer Sozialversicherungen (BVS), Bern. `https://tinyurl.com/altersvorsorge2020-female`

Figure 5: Effects of reform elements on support for the pension reform by party affiliation



Note: Estimates shown are changes in probability of supporting a pension reform that includes the specific level of a reform element, compared to a reform that includes the status quo. Bars represent 95 percent confidence intervals.

However, left-wing voters were indeed more skeptical of the more drastic retrenchment measures discussed: the difference among left-wing voters and the moderate right was significant regarding unbalanced benefit cuts and increasing the retirement age to 67.

In terms of compensation, we see exactly the pattern we would expect. Subsidies for the early retirement of low-wage earners (targeting), better occupational pensions for women (recalibration), and moderately higher revenues via VAT significantly and strongly increased the support for the entire reform package among left-wing voters (by 5–9 percentage points), whereas they did not among voters of either the far or the moderate right. Left-wing voters clearly respond more positively to compensations via targeting and recalibration than right-wing voters do. Hence, compensations make retrenchment acceptable among left-wing voters: 30.2 percent of all reform packages that included pension cutbacks still received majority support among those left voters who explicitly rejected precisely this measure in our survey (in a direct question about these cutbacks).

In terms of a comparative assessment of the effectiveness of different compensation

Table 4: Effects of different compensation strategies on reform support among subgroups of respondents

	Targeting (subsidize early retirement for lower-income earners)	**Recalibration** (extend access to second pillar pensions for women)	**Increased revenues** (increase VAT by 1.5pp)
Full sample	0.012 [0.033 -0.009]	0.045 [0.062 0.028]	0.008 [0.030 -0.014]
Lower-income respondents	0.057 [0.112 0.001]	0.081 [0.131 0.032]	-0.002 [0.064 -0.067]
Women	0.029 [0.063 -0.005]	0.061 [0.090 0.033]	-0.005 [0.031 -0.042]
Left-wing voters	0.074 [0.113 0.035]	0.087 [0.121 0.053]	0.052 [0.092 0.012]

strategies, Table 4 provides the average marginal component effects of the different compensating elements for the various subgroups we have focused on. The coefficients in the cells are simply the coefficients shown in the above figures (95 percent confidence interval below). Regarding which compensation strategy works among which groups, two findings are particularly noteworthy: first, recalibration is most effective in generating support for the overall reform package among all subgroups. In other words: it is the most important compensating factor. Second, partisanship is a more forceful predictor of the effectiveness of compensations than socioeconomic conditions since all three compensation strategies have the most traction among left-wing voters.

5 Robustness Checks

Conjoint surveys are generally cognitively demanding for respondents because they are asked to evaluate several elements in combination and beyond that to compare one combination of elements to another combination. Hence, there are several concerns regarding the validity and robustness of our findings that we want to address directly. In terms of internal validity, one question is whether conjoint evaluations are answered randomly when respondents become tired of the complex task. In addition, the specific formulation of reform options (i.e., values) may affect how respondents understand and evaluate the

packages. On top of this, problems of sampling and survey design (we used a CATI recruitment of respondents) may affect the internal validity of the findings. When it comes to concerns about external validity, the choice and specification of the reform elements (and their levels) can be questioned. Since we have one particular instance of a targeting or recalibrating reform element, can we draw conclusions about the effectiveness of recalibration in general or only about this specific kind of recalibration? Furthermore, we fielded the survey at a specific point in time in terms of the public debate in the media about the ongoing reform. The relative importance respondents attributed to the different reform elements may therefore have been situational and reflected a snapshot of the debate as covered by the media. Finally, one may wonder if the findings only hold for highly sophisticated, well-informed citizens (which raises concerns about both internal and external validity).

We addressed these legitimate concerns via a series of robustness checks. First, we addressed concerns about the effect of respondent fatigue by estimating the key findings on the basis of the first two conjoint comparisons only. The findings are entirely consistent (see Appendix C in the supplementary information), indicating that those respondents who engaged with the task maintained their attention for all five conjoint pairs. This positive finding is also confirmed by an evaluation of time stamps (the time respondents spent on each pairwise comparison) and of dropouts during the conjoint task. Both evaluations show that respondents engaged in all five comparisons seriously and that once they decided to engage with the task they followed through. Only 5.4 percent of the respondents dropped out by the start of the conjoint task, which is the first task in the overall survey.

Second, we repeated a very similar, but not identical survey, about the same pension reform with a new sample about one year after the first fieldwork was completed. This second survey differed in the following respects: (a) In terms of sampling strategy, for the first survey we recruited respondents based on the register of Swiss landline phone numbers via CATI but for the second survey the sample was based on the official statistical register of the Swiss population and respondents were invited via letter. (b) We changed

one of the six reform elements of the survey since the targeting element of the first government proposal (subsidizing early retirement) was changed in the parliamentary debates to an alternative targeting element (increasing basic pensions by a certain amount per month). This allowed us to test the robustness of the component effect depending on the particular kind of targeting compensation. (c) We slightly adapted the formulation of the levels to reflect the evolving policy proposal. This adaptation allowed us to check the robustness of the findings to slight changes in the formulation of levels. (d) Since the second survey was fielded more than a year after the first one, the context in terms of public debate on the reform had changed. While the reform proposal was not yet in parliament in 2015 (hence the public debate in the media had just started), it was more heavily discussed in the media in mid-2016 (i.e., at the time of the second survey) because the process was then in the middle of parliamentary debates. Thus, we can evaluate whether our findings are robust in light of changes in the public opinion context.

Appendix E provides information about the findings of this second survey. The overall picture is very clear: the findings are robust to the changes in sampling, survey design, and context. Comparing the findings of the two surveys, we see an almost identical pattern: retrenchment of second pillar pensions is highly unpopular, especially if it is not balanced with higher contribution payments. Also, a general increase in the age of retirement and restrictions of widows' pensions negatively affect support for the overall package. The finding on recalibrating reform elements is also robust: extending eligibility to low-income and part-time workers significantly increases popular support for the reform (the coefficient for extending eligibility for part-time workers alone is also positive, but not significant). We also confirm that women and left-wing voters support recalibration more strongly than men and right-wing voters. Finally, as in the first survey, increasing revenues through VAT increases support only among left-wing voters. These findings are robust to slightly different formulations of the levels.

The only difference we found concerns targeting compensation. While targeting compensation (subsidizing early retirement for lower-income earners) increased support for the reform in 2015 (first wave) among left-wing voters but not among any other group, the

targeting of compensation included in the 2016 is more effective (and much more generous). In 2016, the upper chamber of parliament proposed to increase basic pension levels by a fixed amount per month for all recipients. Despite being a universalistic measure, this change would benefit lower-income earners most (which is why we still qualify it as targeting) since the basic pensions are highly redistributive. This compensation measure increases support significantly in the overall sample, among left-wing voters, women, and lower-income earners. These findings suggests that targeting may not be a less effective compensation strategy than recalibration overall, but its effect depends on the specific measures proposed.

Finally, Appendix F explains how we tested whether the results differ between unsophisticated and sophisticated voters. We define sophisticated voters as those with tertiary education. This item measures above all the "ability"-aspect of political sophistication, which is distinct from, but related to the motivational and informational aspects of sophistication (Luskin, 1990). Since we are interested here in the performance of respondents in the complex conjoint task, the ability-dimension of sophistication is relevant. First, we wanted to see if respondents with different levels of education were equally consistent in performing the conjoint task. For this, we estimated the AMCEs for three different splits of the five paired comparisons and for low- and highly educated respondents separately. We then correlated the AMCEs of the first two (or three) comparisons with the AMCEs of the second set of comparisons. Since the distribution of respondents with and without tertiary education is highly skewed in our sample (about one-third with tertiary education and two-thirds without), we drew 1,000 repeated random samples and compared the average of computed correlation measures. All comparisons showed that less sophisticated respondents had performed the task just as consistently as more sophisticated respondents. Beyond this, Table 6 in Appendix D also shows that the findings are robust for a simple control for tertiary education. Moreover, we tested whether more and less educated respondents had answered the conjoint tasks consistently with direct, simple survey questions about specific reform elements. Our survey contains three direct questions about reform elements that correspond directly to conjoint levels. Hence, we correlated

the individual-level answers respondents gave on these direct questions (likert scales, 1-4) with the individual-level AMCEs, estimated on the basis of the 10 conjoint packages each respondent had evaluated, again for 1,000 randomly drawn samples of high- and low-educated respondents. Several estimators of correlation (see Appendix F.2) clearly indicate that while the more highly educated respondents are—as expected—somewhat more coherent in their answers, the coherence between the two groups never differs significantly. Based on all these tests, we feel confident that the conjoint task was understood and performed well by both more sophisticated and less sophisticated respondents.

6 Conclusions

The question of under what conditions welfare states can be reformed in times of austerity has become crucial in research on welfare politics. Compensation as a strategy to foster support for reform is one of the key explanations. Compensation presents (parts of) the opponents with a trade-off: while they may reject certain elements of a reform, they are interested in securing other elements. Conjoint analysis is able to capture precisely the microlevel mechanisms of evaluating trade-offs. It allows us to answer empirically a range of crucial research questions regarding the politics of compensation: Are there policy reform packages that receive majority support when combining retrenchment and compensation? Which compensation strategies are most effective? And what factors determine the effectiveness of different kinds of compensation among particular social groups?

On the basis of our original survey of 1,873 Swiss voters, we have shown that the retrenchment of existing benefit levels is indeed highly unpopular. About 60 percent of respondents consider lower pension benefits unacceptable. Even when retrenchment is embedded in a wider, more balanced reform package, it has a clear negative effect on the acceptance of the reform. More concretely, lowering pension benefit levels in the occupational pension pillar reduces the probability of a reform being accepted in a direct democratic referendum by 5 to 10 percentage points (the AMCE), compared to a

reform proposal that retains the status quo in terms of benefit levels. This reflects the "cost" in terms of public opinion that compensation needs to counter-balance. Different compensation strategies are able to counterbalance this negative effect on reform support. Adapting the architecture of the pension system to new social risks (recalibration) has the strongest and most consistent positive effect on reform support: it increases the chances of a reform being accepted significantly by about 5 percentage points in the full sample and by about 6 to 8 percentage points among the main beneficiaries (women and lower-income earners). Expansion for lower-income earners (targeting) has a more variable effect: when it comes in the form of subsidies for early retirement, it does not increase the support for a reform significantly among the entire population, but only among the main beneficiaries (low-income earners). Targeting is more effective when designed more widely (i.e., in the form of generally higher basic pensions).

However, our findings also show that material conditions only take us so far in explaining citizens' responses to compensation. Partisanship has stronger effects in predicting citizens' responses. Compensation has clear and positive effects among voters of the left: both targeting and recalibration increase the probability of a reform package being accepted by about 7 and 9 percentage points respectively. Increasing revenues (from VAT) adds about 5 percentage points of support. This is an important finding, given the fact that it is usually the Left that organizes opposition against pension cutbacks. It implies that if the compensation elements in the reform proposal divide the left-wing opposition, the chances of the reform increase.

These findings obviously have important implications for the reform capacity of mature welfare states in an era of permanent financial constraint. Not only do they show that broad reform packages are more likely to gather sufficient support than narrow reform proposals, but they also confirm that structural and institutional constraints are not deterministic. Rather, agency and politics matter: governments do have ample room for maneuvering to tailor compensations to the relevant opposition groups.

These implications matter beyond the case of pension reform in Switzerland. In terms of scope conditions and external validity, we need to reflect on the extent to which these

findings travel beyond the country and policy field. Regarding the country-context, we contend that our results do bear relevance for other power-sharing democracies for several reasons: first, the scenario of a reform proposal that cuts back on pension expenditure while such cutbacks are massively unpopular in the public is highly similar to reform contexts across mature welfare states. The same goes for the political Left as the most likely and capable actor of mobilizing political opposition against such reform. Furthermore, Switzerland provides an ideal case to study the effectiveness of compensation, because major pension reforms are usually subject to a direct democratic popular vote. Over the past 25 years, no less than 7 such referenda dealt with pension reform. This provides for a highly realistic setting for our survey experiment, as the task we ask respondents to perform is very close to a real-world task. We also contend that this institutional specificity does not generally limit the relevance of our findings beyond Switzerland (even though other countries may not usually have referenda on pension reforms), because citizens can and do protest pension cutbacks via other channels, such as demonstrations or strikes. We also made a specific effort (via media analysis and expert validation) to ensure that the reform elements we included in the conjoint experiments were all highly visible, politicized and understood in the public. Hence, we are confident that they would also give rise to public debates in other contexts. Various pre-tests and robustness checks confirmed that all respondents, not only the most well-informed or sophisticated ones, did understand the task at hand. Hence, we feel confident that our findings do bear relevance to pension reform beyond Switzerland.

A second potential scope condition relates to the perception of fiscal constraint and a trade-off scenario among respondents, i.e. whether they agree with the basic assumption that pure expansion is not a viable reform option. If only those respondents who subscribe to such a fiscal constraint were susceptible to compensation, this would limit the validity of our findings to a sub-group of the population. This sub-group is small in Switzerland (about 12% only), but may be larger elsewhere. We were able to test this condition empirically and found that compensation enhances support for the reform even among those citizens who think that retrenchment pressure is not high. Hence, we feel confident

that our findings matter also in contexts where the basic need for a reform is more contested than in Switzerland.

Finally, one may ask if our findings bear any relevance beyond old age pensions as a policy field. We do acknowledge that pension policy is the prime example of a field in which cutbacks are broadly and strongly contested and where compensation is both necessary and possible. Where risks are more concentrated and hence cutbacks affect smaller segments of the population (e.g. unemployment reform), compensation may be both less necessary politically and less effective, because fewer people perceive a direct material interest in them. And in areas where cutbacks are less directly visible to the public (possibly in health), compensation may also be less relevant. However, this is an empirical question that we could not test in the scope of this study. Similarly, the relative effectiveness of recalibration, targeting and increasing revenues as alternative compensation strategies may depend both on the context in which they resonate, and on the precise design. We found that targeting is more effective when it affects basic pension generosity than when it addresses specific needs. Hence, further research may add to our understanding of the conditions that shape the effectiveness of different compensation strategies.

Beyond the specific findings of this study on compensation and reform capacity, we want to emphasize the potential and usefulness of conjoint analysis for the study of current welfare politics. We see four main assets of conjoint analysis. First, almost all current research on welfare state change relies on theoretical arguments involving individual-level preferences and public opinion as key mechanisms in political dynamics. However, we know from survey research that when asked directly, most people tend to support generous welfare spending in all areas, just as they support low tax levels (which is why we have such skewed distributions in general survey questions on the welfare state). It is hard to add constraints to such questions in a way that reveals narrower preferences and increases variance. Conjoint analysis allows for modeling realistic constraints directly.

Second, introducing such realistic constraints is particularly important in current welfare state research since most of our theoretical arguments actually rely on assumptions

regarding policy priorities rather than policy position. Almost all respondents reject pension retrenchment, but this retrenchment does not have the same importance for all respondents. Priorities have become highly relevant politically because the context of contemporary welfare politics resembles a zero-sum distributive game in which gains for some social groups come at the expense of other groups. In this context, conjoint analysis provides us with a tool to conceptualize and measure the relative importance of multiple desirable goals. It therefore has tremendous potential for applications in welfare state research way beyond the question of retrenchment and reform capacity.

Third, conjoint analysis is an experimental survey technique. Its strength therefore naturally lies more in internal validity than in external validity. However, we have reason to think that conjoint analysis provides more external validity than traditional survey experiments (Hainmueller et al., 2014, p. 27) because conjoint pairs can be designed in a highly realistic way and may therefore capture actual opinion-formation processes more adequately than traditional surveys that tend to place respondents in a more artificial situation. In the real world, citizens evaluate policy packages when deciding about their support for a reform, a government, or a party. Hence, conjoint surveys may actually achieve more external validity than traditional surveys.

Finally, and because of the realistic decision-making situation conjoint analysis can create for respondents, the policy-relevance of our empirical findings may be higher than in traditional survey research that is unable to capture the complexity and multidimensionality of the decisions at hand.

One of the downsides of conjoint analysis, however, is that the findings may not be very robust since they depend on the precise choice and wording of the elements and levels, and they may depend on the current public debate (i.e., they reflect momentary snapshots of the public debate rather than stable preferences). In our study, we were able to contain these problems via a robustness survey testing similar but not identical reform elements/levels, which was fielded roughly one year after the first data collection. However, future researchers may want to go beyond this by theorizing and explicitly testing the context effect of public debates on the relative importance citizens attribute

to particular compensations via either longitudinal or comparative designs.

References

Ballard-Rosa, Cameron, Martin, Lucy, & Scheve, Kenneth (2017). The structure of American income tax policy preferences. *The Journal of Politics*, *79*(1), 1–16.

Bechtel, Michael, Genovese, Federica, & Scheve, Kenneth F. (2017). Interests, Norms, and Support for the Provision of Global Public Goods: The Case of Climate Cooperation. *British Journal of Political Science*, (Forthcoming).

Bechtel, Michael M., Hainmueller, Jens, & Margalit, Yotam (2014). Preferences for international redistribution: The divide over the Eurozone bailouts. *American Journal of Political Science*, *58*(4), 835–856.

Bertrand, Marianne & Mullainathan, Sendhil (2001). Do People Mean What They Say? Implications For Subjective Survey Data. SSRN Electronic Journal. `https://papers.ssrn.com/sol3/papers.cfm?abstract_id=260131`.

Bonoli, Giuliano (2000). *The politics of pension reform: institutions and policy change in Western Europe*. Cambridge: Cambridge University Press.

Bonoli, Giuliano (2005). The politics of the new social policies: providing coverage against new social risks in mature welfare states. *Policy & Politics*, *33*(3), 431–449.

Bonoli, Giuliano & Natali, David (Eds.). (2012). *The Politics of the New Welfare State*. Oxford, New York: Oxford University Press.

Bornschier, Simon (2015). The new cultural conflict, polarization, and representation in the Swiss party system, 1975–2011. *Swiss Political Science Review*, *21*(4), 680–701.

Brooks, Clem & Manza, Jeff (2008). *Why welfare states persist: the importance of public opinion in democracies*. Chicago: University of Chicago Press.

Busemeyer, Marius R. & Garritzmann, Julian L. (2017). Public opinion on policy and budgetary trade-offs in European welfare states: evidence from a new comparative survey. *Journal of European Public Policy*, *24*(6), 871–889.

Busemeyer, Marius R. & Neimanns, Erik (2017). Conflictive preferences towards social investments and transfers in mature welfare states: The cases of unemployment benefits and childcare provision. *Journal of European Social Policy*, *27*(3), 229–246.

Cox, Robert Henry (2001). The Social Construction of an Imperative: Why Welfare Reform Happened in Denmark and the Netherlands but Not in Germany. *World Politics*, *53*(3), 463–498.

Esping-Andersen, Gøsta (1999). *Social foundations of postindustrial economies*. Oxford: Oxford University Press.

Gallego, Aina & Marx, Paul (2017). Multi-dimensional preferences for labour market reforms: a conjoint experiment. *Journal of European Public Policy*, *24*(7), 1027–1047.

Giger, Nathalie & Nelson, Moira (2011). The electoral consequences of welfare state retrenchment: Blame avoidance or credit claiming in the era of permanent austerity? *European Journal of Political Research*, *50*(1), 1–23.

Giger, Nathalie & Nelson, Moira (2013). The welfare state or the economy? Preferences, constituencies, and strategies for retrenchment. *European Sociological Review*, *29*(5), 1083–1094.

Gingrich, Jane (2014). Visibility, values, and voters: The informational role of the welfare state. *The Journal of Politics*, *76*(2), 565–580.

Green, Paul E., Krieger, Abba M., & Wind, Yoram (2001). Thirty years of conjoint analysis: Reflections and prospects. *Interfaces*, *31*, 56–73.

Hacker, Jacob S. (2004). Privatizing risk without privatizing the welfare state: The hidden politics of social policy retrenchment in the United States. *American Political Science Review*, *98*(02), 243–260.

Hainmueller, Jens, Hangartner, Dominik, & Yamamoto, Teppei (2015). Validating vignette and conjoint survey experiments against real-world behavior. *Proceedings of the National Academy of Sciences*, *112*(8), 2395–2400.

Hainmueller, Jens & Hopkins, Daniel J. (2014). The Hidden American Immigration Consensus: A Conjoint Analysis of Attitudes toward Immigrants. *American Journal of Political Science*, *59*(3), 529–548.

Hainmueller, J., Hopkins, D. J., & Yamamoto, T. (2014). Causal Inference in Conjoint Analysis: Understanding Multidimensional Choices via Stated Preference Experiments. *Political Analysis*, *22*(1), 1–30.

Häusermann, Silja (2010). *The Politics of Welfare State Reform in Continental Europe: Modernization in Hard Times*. Cambridge: Cambridge University Press.

Hemerijck, Anton (2013). *Changing welfare states*. Oxford: Oxford University Press.

Horiuchi, Yusaku, Smith, Daniel M., & Yamamoto, Teppei (2015). Measuring voters' multidimensional policy preferences with conjoint analysis: Application to Japan's 2014 election. SSRN Electronic Journal. `http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2627907`.

Huber, Evelyne & Stephens, John D. (2001). *Development and crisis of the welfare state: parties and policies in global markets*. Chicago: The University of Chicago Press.

Huber, Evelyne & Stephens, John D. (2015). Postindustrial Social Policy. In P. Beramendi, S. Häusermann, H. Kitschelt, & H. Kriesi (Eds.), *The politics of advanced capitalism* (pp. 112–132). New York: Cambridge University Press.

Immergut, Ellen M., Anderson, Karen M., & Schulze, Isabelle (2007). *The Handbook of West European Pension Politics*. OUP Oxford.

Jensen, Carsten (2012). Labour market-versus life course-related social policies: understanding cross-programme differences. *Journal of European Public Policy*, *19*(2), 275–291.

Kitschelt, Herbert (2001). Partisan Competition and Welfare State Retrenchment: When Do Politicians Choose Unpopular Policies? In P. Pierson (Ed.), *The new politics of the welfare state* (pp. 265–302). Oxford: Oxford University Press.

Knotz, Carlo & Lindvall, Johannes (2015). Coalitions and compensation: The case of unemployment benefit duration. *Comparative Political Studies*, *48*(5), 586–615.

Kuhnle, Stein (2002). *Survival of the European welfare state*. Routledge/ECPR studies in European political science. London: Routledge.

Levy, Jonah D. (1999). Vice into virtue? Progressive politics and welfare reform in continental Europe. *Politics & Society*, *27*(2), 239–273.

Lindvall, Johannes (2017). Economic Downturns and Political Competition since the 1870s. *The Journal of Politics*, *79*(4), 1302–1314.

Luskin, Robert C. (1990). Explaining political sophistication. *Political Behavior*, *12*(4), 331–361.

Mutz, Diana C. (1998). *Impersonal Influence: How Perceptions of Mass Collectives Affect Political Attitudes*. Cambridge u.a.: Cambridge University Press: Cambridge University Press.

Natali, David & Rhodes, Martin (2004). Trade-offs and veto players: reforming pensions in France and Italy. *French Politics*, *2*(1), 1–23.

Palier, Bruno (2010). *A long goodbye to Bismarck?: the politics of welfare reforms in continental Europe*. Changing welfare states. Amsterdam: Amsterdam University Press.

Pierson, Paul (2000). Increasing returns, path dependence, and the study of politics. *American Political Science Review*, *94*(02), 251–267.

Pierson, Paul (2001a). Coping with Permanent Austerity. Welfare State Restructuring in Affluent Democracies. In P. Pierson (Ed.), *The new politics of the welfare state* (pp. 410–456). Oxford: Oxford University Press.

Pierson, Paul (2001b). *The new politics of the welfare state*. Oxford: Oxford University Press.

Rehm, Philipp, Hacker, Jacob S., & Schlesinger, Mark (2012). Insecure alliances: Risk, inequality, and support for the welfare state. *American Political Science Review*, *106*(02), 386–406.

Rehm, Philipp Benjamin (2016). *Risk inequality and welfare states: social policy preferences, development, and dynamics* (Frist published ed.). Cambridge studies in comparative politics. New York NY: Cambridge University Press.

Rhodes, Martin (2001). The Political Economy of Social Pacts: Competitive Corporatism and European Welfare Reform. In P. Pierson (Ed.), *The new politics of the welfare state*. Oxford: Oxford University Press.

Vail, Mark (2010). *Recasting welfare capitalism: economic adjustment in contemporary France and Germany*. Philadelphia: Temple University Press.

Van Oorschot, Wim (2006). Making the difference in social Europe: deservingness perceptions among citizens of European welfare states. *Journal of European social policy*, *16*(1), 23–42.

Vis, Barbara (2016). Taking Stock of the Comparative Literature on the Role of Blame Avoidance Strategies in Social Policy Reform. *Journal of Comparative Policy Analysis: Research and Practice*, *18*(2), 122–137.

Wallander, Lisa (2009). 25 years of factorial surveys in sociology: A review. *Social Science Research*, *38*(3), 505–520.

Supplementary Information

A Conjoint pairwise comparison - screenshot of the online survey

Figure 1: Conjoint pairwise comparison - screenshot of the online survey

	Reformpaket 1	Reformpaket 2
Eintrittschwelle in Pensionskasse	Unverändert: keine Ausweitung des Zugangs zu Renten aus Pensionskassen.	Erweiterter Zugang zu Renten aus Pensionskassen für Personen mit tieferem Einkommen und Teilzeitbeschäftigte.
Witwenrente	Wird schrittweise abgeschafft.	Keine Kürzung.
Mehreinnahmen für die AHV	Erhöhung der Mehrwertsteuer um maximal 1.5 Prozentpunkte.	Erhöhung der Mehrwertsteuer um maximal 3 Prozentpunkte.
Umwandlungssatz Pensionskasse	Senkung. Ausgleich durch höheres Altersguthaben.	Senkung. Ausgleich durch höheres Altersguthaben.
Flexibles Rentenalter	Wie bisher: Frühpensionierung möglich, jedoch mit gekürzter Rente.	Frühpensionierung möglich. Neu: finanzielle Abfederung für Personen mit tiefem Einkommen.
Rentenalter	Schrittweise Erhöhung auf 67 Jahre für Männer und Frauen.	Bleibt stabil: 64 für Frauen, 65 für Männer.

Falls Sie noch einmal nachlesen möchten, worum es bei den Reformelementen geht, klicken Sie bitte hier.
Um zurück zur Umfrage zu gelangen, können Sie einfach oben wieder in den gewünschten Tab klicken.

Welches Reformpaket bevorzugen Sie?

○ Reformpaket 1 ○ Reformpaket 2

Wenn Sie über **Reformpaket 1** abstimmen müssten, würden Sie der Reform zustimmen oder sie ablehnen?

○ sicher zustimmen ○ eher zustimmen ○ eher ablehnen ○ sicher ablehnen ○ weiss nicht / keine Antwort

Wenn Sie über **Reformpaket 2** abstimmen müssten, würden Sie der Reform zustimmen oder sie ablehnen?

○ sicher zustimmen ○ eher zustimmen ○ eher ablehnen ○ sicher ablehnen ○ weiss nicht / keine Antwort

B Expenditures

Table 1: Expenditures for different reform elements

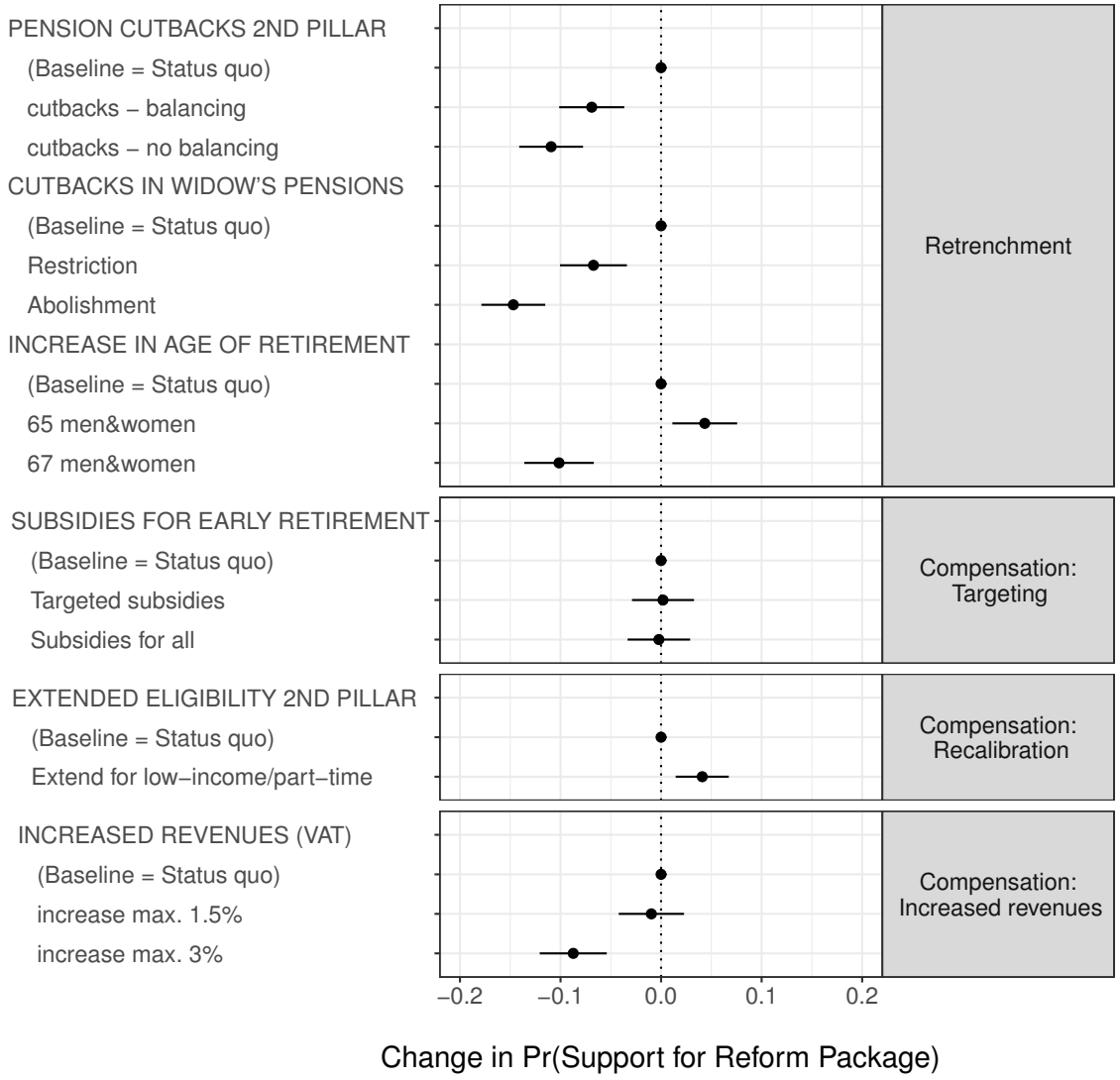
Reform element	Levels	Change in expenditures	Source of the Estimate
Pension cutbacks	1: status quo	0	
2nd pillar	2: Cutbacks. Balanced with higher contribution payments	-1406 mn/year	BBI 2014
	3: Cutbacks. No balancing	-4116 mn/year	BBI 2014
Cutbacks in	1: status quo	0	
widows' pensions	2: Restriction of eligibility	-359 mn/year	BBI 2014
	3: Stepwise abolishment	-960 mn/year	BSV 2016
Increase in age of	1: status quo	0	
retirement	2: 65 for men and women	-1114 mn/year	BBI 2014
	3: Stepwise increase for both men & women to 67	-4700 mn/year	BBI 2014
Subsidies for	1: status quo	0	
early retirement	2: Subsidies for lower-income earners	+390 mn/year	BBI 2014
Extended elibility	1: status quo	0	
2nd pillar	2: Extend access for people with lower incomes and part-time workers	+400 mn/year	BBI 2014
Increased revenues	1: status quo	0	
(VAT)	2: Increase by max. 1.5 pp	0 (revenue increase of 3600 mn/year)	BSV 2014
	3: Increase by max. 3 pp	0 (revenue increase of 7200 mn/year)	BSV 2014

Sources:
BBl 2014: Botschaft zur Reform der Altersvorsorge 2020 vom 19. November 2014 (Official bill proposal by the Federal Government to the Parliament), Bundesblatt, reference number 14.088. https://www.admin.ch/gov/de/start/bundesrecht/bundesblatt.html.
BSV 2016: Bundesamt für Sozialversicherungen. Schweizerische Sozialversicherungsstatistik 2016 (Official social insurance statistics of 2016). Bern. www.bsv.admin.ch/bsv/de/home/sozialversicherungen/ueberblick/grsv/statistik.html.
BSV 2014: Bundesamt für Sozialversicherungen. Zusammenfassung der Vernehmlassungsergebnisse (Synthesis of the consultation procedure), Bern. https://www.bsv.admin.ch/bsv/de/home/sozialversicherungen/ahv/reformen-revisionen/altersvorsorge2020/dokumentation.html

C Robustness checks I

First and second pair only

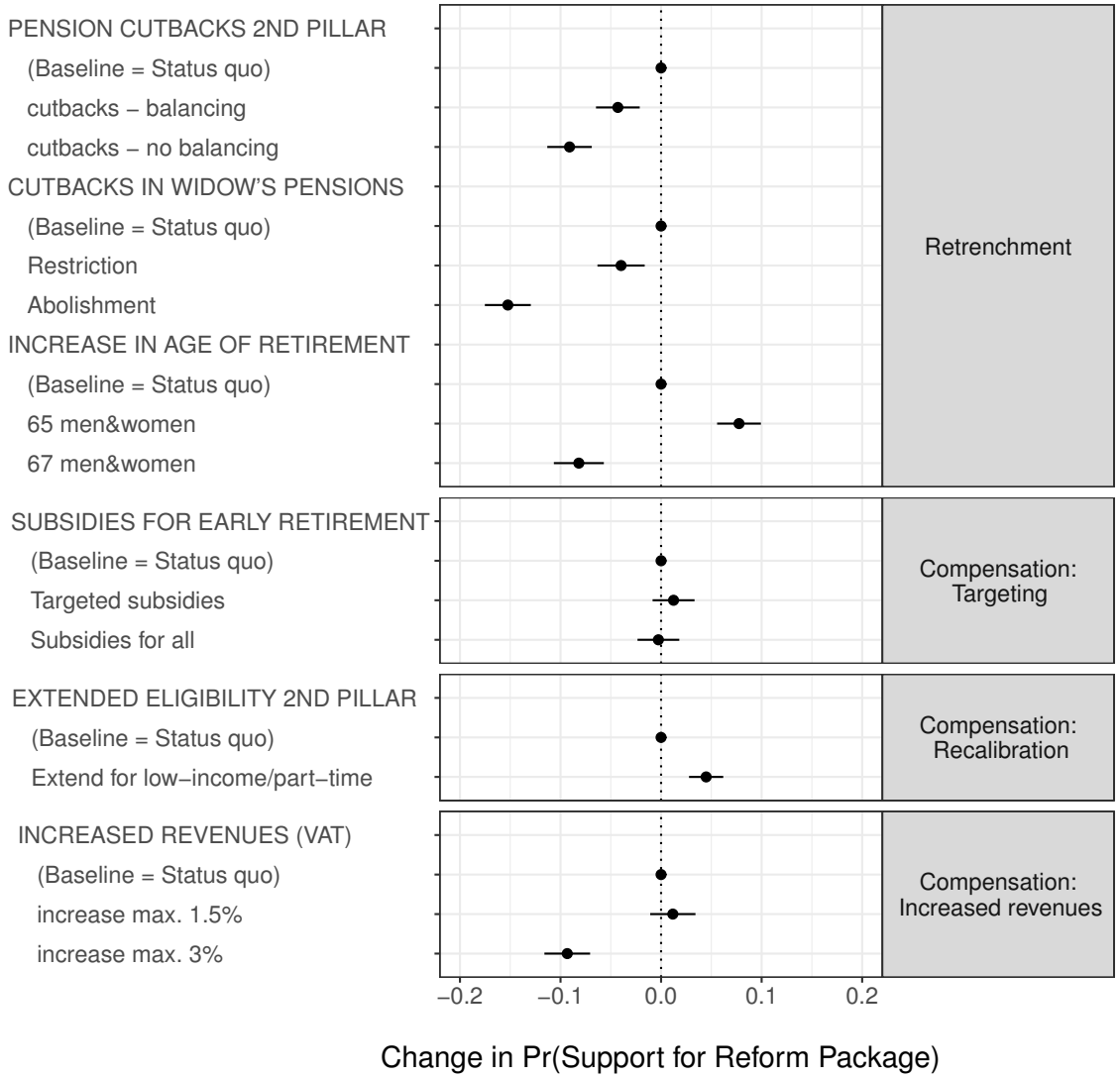
Figure 2: Effects of reform elements on support for the pension reform package, only first two conjoint comparisons



Note: Findings based on a sub-sample of hypothetical packages that includes only the first two (out of five) conjoint comparisons. N = 7492 packages.

Excluding packages with only status quo on all retrenchment elements

Figure 3: Effects of reform elements on support for the pension reform package, only packages with cutbacks on at least one of the retrenchment-elements



Note: Findings based on a sub-sample of hypothetical packages that includes only those hypothetical packages that contain cutbacks (i.e. not status quo) on at least one of the 3 retrenchment reform elements. Of 18730 packages, 683 contained status quo on all three retrenchment-elements.

D Robustness checks II: Controls

Table 2: Full models, control for education

	Model 1	Model 2 Control variables
(Intercept)	0.620	0.621
	[0.590; 0.649]	[0.591; 0.650]
Conversion_Ratecutbacks - balancing	−0.046	−0.045
	[−0.067; −0.025]	[−0.066; −0.024]
Conversion_Ratecutbacks - no balancing	−0.094	−0.094
	[−0.115; −0.073]	[−0.115; −0.072]
Widows_PensionsRestriction	−0.042	−0.042
	[−0.065; −0.020]	[−0.065; −0.019]
Widows_PensionsAbolishment	−0.155	−0.155
	[−0.177; −0.133]	[−0.177; −0.133]
Retirement_Age65 men&women	0.075	0.075
	[0.054; 0.096]	[0.054; 0.096]
Retirement_Age67 men&women	−0.085	−0.084
	[−0.109; −0.060]	[−0.108; −0.059]
Early_RetirementTargeted subsidies	0.012	0.012
	[−0.009; 0.033]	[−0.009; 0.033]
Early_RetirementSubsidies for all	−0.003	−0.004
	[−0.024; 0.017]	[−0.024; 0.017]
Eligibility_2nd_PillarExtend for low-income/part-time	0.045	0.045
	[0.028; 0.062]	[0.028; 0.061]
VATincrease max. 1.5%	0.008	0.008
	[−0.014; 0.030]	[−0.014; 0.030]
VATincrease max. 3%	−0.095	−0.095
	[−0.117; −0.072]	[−0.117; −0.073]
High education		−0.004
		[−0.008; 0.000]
Deviance	4441.562	4422.415
Dispersion	0.237	0.237
Num. obs.	18730	18660

Table 3: Models low and high income, control for ideology, age, gender

	Model 1	Model 2 Control variables
(Intercept)	0.665	0.669
	[0.621; 0.708]	[0.623; 0.714]
Conversion_Ratecutbacks - balancing	−0.045	−0.045
	[−0.076; −0.015]	[−0.076; −0.015]
Conversion_Ratecutbacks - no balancing	−0.113	−0.113
	[−0.145; −0.082]	[−0.145; −0.082]
Widows_PensionsRestriction	−0.060	−0.060
	[−0.094; −0.025]	[−0.095; −0.026]
Widows_PensionsAbolishment	−0.173	−0.173
	[−0.206; −0.140]	[−0.206; −0.140]
Retirement_Age65 men&women	0.057	0.057
	[0.025; 0.088]	[0.025; 0.088]
Retirement_Age67 men&women	−0.114	−0.114
	[−0.150; −0.078]	[−0.150; −0.078]
Early_RetirementTargeted subsidies	0.003	0.003
	[−0.029; 0.034]	[−0.028; 0.035]
Early_RetirementSubsidies for all	−0.014	−0.013
	[−0.042; 0.015]	[−0.042; 0.015]
Eligibility_2nd_PillarExtend for low-income/part-time	0.038	0.039
	[0.014; 0.063]	[0.014; 0.063]
VATincrease max. 1.5%	0.009	0.009
	[−0.024; 0.041]	[−0.024; 0.042]
VATincrease max. 3%	−0.103	−0.103
	[−0.136; −0.071]	[−0.136; −0.071]
Low income	−0.053	−0.052
	[−0.144; 0.038]	[−0.143; 0.039]
Conversion_Ratecutbacks - balancing: low income	−0.029	−0.028
	[−0.098; 0.041]	[−0.098; 0.041]
Conversion_Ratecutbacks - no balancing: low income	0.024	0.024
	[−0.043; 0.091]	[−0.044; 0.091]
Widows_PensionsRestriction: low income	0.003	0.004
	[−0.070; 0.076]	[−0.070; 0.077]
Widows_PensionsAbolishment: low income	−0.023	−0.023
	[−0.089; 0.042]	[−0.089; 0.043]
Retirement_Age65 men&women: low income	0.019	0.019
	[−0.053; 0.090]	[−0.053; 0.090]
Retirement_Age67 men&women: low income	0.007	0.007
	[−0.075; 0.089]	[−0.075; 0.089]
Early_RetirementTargeted subsidies: low income	0.054	0.053
	[−0.010; 0.118]	[−0.011; 0.117]
Early_RetirementSubsidies for all: low income	0.046	0.046
	[−0.021; 0.113]	[−0.021; 0.113]
Eligibility_2nd_PillarExtend for low-income/part-time: low income	0.043	0.043
	[−0.012; 0.098]	[−0.012; 0.098]
VATincrease max. 1.5%: low income	−0.011	−0.011
	[−0.084; 0.062]	[−0.084; 0.062]
VATincrease max. 3%: low income	0.009	0.009
	[−0.065; 0.083]	[−0.066; 0.083]
Ideology: left		−0.001
		[−0.011; 0.009]
Ideology: moderate right		−0.001
		[−0.011; 0.009]
Ideology: other		−0.006
		[−0.015; 0.003]
Age		−0.000
		[−0.000; 0.000]
Female		−0.004
		[−0.010; 0.002]
Deviance	2253.421	2253.319
Dispersion	0.234	0.234
Num. obs.	9640	9640

Table 4: Models gender, control for ideology, age, income

	Model 1	Model 2 Control variables
(Intercept)	0.666	0.662
	[0.618; 0.714]	[0.608; 0.717]
Conversion_Ratecutbacks - balancing	−0.077	−0.076
	[−0.115; −0.039]	[−0.115; −0.036]
Conversion_Ratecutbacks - no balancing	−0.119	−0.128
	[−0.158; −0.080]	[−0.169; −0.087]
Widows_PensionsRestriction	−0.025	−0.032
	[−0.065; 0.015]	[−0.075; 0.011]
Widows_PensionsAbolishment	−0.149	−0.151
	[−0.188; −0.110]	[−0.192; −0.109]
Retirement_Age65 men&women	0.091	0.095
	[0.056; 0.127]	[0.057; 0.134]
Retirement_Age67 men&women	−0.085	−0.080
	[−0.127; −0.043]	[−0.125; −0.034]
Early_RetirementTargeted subsidies	0.002	0.001
	[−0.035; 0.040]	[−0.038; 0.040]
Early_RetirementSubsidies for all	−0.013	−0.013
	[−0.048; 0.023]	[−0.051; 0.024]
Eligibility_2nd_PillarExtend for low-income/part-time	0.019	0.026
	[−0.009; 0.047]	[−0.004; 0.057]
VATincrease max. 1.5%	−0.008	0.002
	[−0.047; 0.031]	[−0.039; 0.044]
VATincrease max. 3%	−0.136	−0.123
	[−0.175; −0.097]	[−0.165; −0.080]
Female	−0.017	−0.008
	[−0.087; 0.052]	[−0.084; 0.068]
Conversion_Ratecutbacks - balancing: female	0.046	0.049
	[−0.005; 0.097]	[−0.006; 0.105]
Conversion_Ratecutbacks - no balancing: female	0.041	0.039
	[−0.011; 0.093]	[−0.016; 0.095]
Widows_PensionsRestriction: female	−0.053	−0.054
	[−0.108; 0.002]	[−0.114; 0.007]
Widows_PensionsAbolishment: female	−0.050	−0.052
	[−0.102; 0.002]	[−0.109; 0.005]
Retirement_Age65 men&women: female	−0.068	−0.066
	[−0.119; −0.017]	[−0.122; −0.010]
Retirement_Age67 men&women: female	−0.062	−0.065
	[−0.120; −0.004]	[−0.129; −0.001]
Early_RetirementTargeted subsidies: female	0.027	0.026
	[−0.024; 0.077]	[−0.029; 0.081]
Early_RetirementSubsidies for all: female	0.037	0.019
	[−0.011; 0.085]	[−0.033; 0.071]
Eligibility_2nd_PillarExtend for low-income/part-time: female	0.043	0.040
	[0.002; 0.083]	[−0.003; 0.084]
VATincrease max. 1.5%: female	0.003	0.007
	[−0.050; 0.056]	[−0.051; 0.066]
VATincrease max. 3%: female	0.050	0.042
	[−0.004; 0.104]	[−0.017; 0.100]
Ideology: left		−0.001
		[−0.011; 0.008]
Ideology: moderate right		−0.002
		[−0.012; 0.009]
Ideology: other		−0.007
		[−0.016; 0.002]
Age		−0.000
		[−0.000; 0.000]
Low income		0.003
		[−0.005; 0.010]
Deviance	2676.822	2249.158
Dispersion	0.234	0.233
Num. obs.	11450	9640

Table 5: Models party affiliation

	Model 1	Model 2 Control variables
(Intercept)	0.582 [0.520; 0.644]	0.557 [0.490; 0.624]
Conversion_Ratecutbacks - balancing	−0.046 [−0.091; −0.002]	−0.040 [−0.088; 0.008]
Conversion_Ratecutbacks - no balancing	−0.053 [−0.099; −0.006]	−0.054 [−0.104; −0.005]
Widows_PensionsRestriction	−0.026 [−0.075; 0.023]	−0.019 [−0.074; 0.035]
Widows_PensionsAbolishment	−0.130 [−0.177; −0.083]	−0.125 [−0.177; −0.073]
Retirement_Age65 men&women	0.101 [0.056; 0.146]	0.110 [0.062; 0.159]
Retirement_Age67 men&women	−0.003 [−0.054; 0.048]	0.013 [−0.041; 0.067]
Early_RetirementTargeted subsidies	−0.019 [−0.066; 0.029]	−0.029 [−0.082; 0.024]
Early_RetirementSubsidies for all	−0.037 [−0.083; 0.008]	−0.039 [−0.089; 0.011]
Eligibility_2nd_PillarExtend for low-income/part-time	0.015 [−0.018; 0.048]	0.021 [−0.015; 0.057]
VATincrease max. 1.5%	0.015 [−0.038; 0.067]	0.033 [−0.023; 0.088]
VATincrease max. 3%	−0.065 [−0.115; −0.015]	−0.055 [−0.111; −0.000]
ideolfarright	0.140 [0.041; 0.240]	0.156 [0.047; 0.264]
ideolleft	−0.024 [−0.108; 0.060]	−0.016 [−0.106; 0.074]
Conversion_Ratecutbacks - balancing:ideolfarright	−0.016 [−0.088; 0.056]	−0.020 [−0.097; 0.058]
Conversion_Ratecutbacks - no balancing:ideolfarright	−0.045 [−0.118; 0.029]	−0.050 [−0.127; 0.028]
Conversion_Ratecutbacks - balancing:ideolleft	0.002 [−0.060; 0.064]	−0.006 [−0.072; 0.060]
Conversion_Ratecutbacks - no balancing:ideolleft	−0.064 [−0.127; −0.001]	−0.066 [−0.132; 0.001]
Widows_PensionsRestriction:ideolfarright	−0.070 [−0.151; 0.011]	−0.078 [−0.166; 0.010]
Widows_PensionsAbolishment:ideolfarright	−0.026 [−0.103; 0.051]	−0.029 [−0.114; 0.055]
Widows_PensionsRestriction:ideolleft	0.011 [−0.053; 0.076]	0.004 [−0.066; 0.074]
Widows_PensionsAbolishment:ideolleft	−0.022 [−0.086; 0.041]	−0.028 [−0.097; 0.041]

Table 6: Models party affiliation,cont.

	Model 1	Model 2 Control variables
Retirement_Age65 men&women:ideolfarright	−0.054	−0.046
	[−0.128; 0.021]	[−0.126; 0.034]
Retirement_Age67 men&women:ideolfarright	−0.084	−0.103
	[−0.168; 0.000]	[−0.195; −0.011]
Retirement_Age65 men&women:ideolleft	−0.037	−0.043
	[−0.097; 0.022]	[−0.106; 0.021]
Retirement_Age67 men&women:ideolleft	−0.132	−0.137
	[−0.202; −0.063]	[−0.210; −0.063]
Early_RetirementTargeted subsidies:ideolfarright	−0.016	−0.020
	[−0.089; 0.057]	[−0.100; 0.060]
Early_RetirementSubsidies for all:ideolfarright	0.060	0.051
	[−0.009; 0.128]	[−0.022; 0.125]
Early_RetirementTargeted subsidies:ideolleft	0.092	0.107
	[0.031; 0.154]	[0.040; 0.174]
Early_RetirementSubsidies for all:ideolleft	0.046	0.052
	[−0.015; 0.107]	[−0.014; 0.117]
Eligibility_2nd_PillarExtend for low-income/part-time:ideolfarright	−0.022	−0.020
	[−0.078; 0.034]	[−0.081; 0.040]
Eligibility_2nd_PillarExtend for low-income/part-time:ideolleft	0.072	0.072
	[0.024; 0.120]	[0.020; 0.123]
VATincrease max. 1.5%:ideolfarright	−0.035	−0.037
	[−0.118; 0.047]	[−0.126; 0.051]
VATincrease max. 3%:ideolfarright	−0.093	−0.099
	[−0.171; −0.016]	[−0.183; −0.014]
VATincrease max. 1.5%:ideolleft	0.037	0.024
	[−0.029; 0.103]	[−0.046; 0.094]
VATincrease max. 3%:ideolleft	0.025	0.027
	[−0.040; 0.089]	[−0.043; 0.097]
female1		−0.002
		[−0.008; 0.005]
age		0.000
		[−0.000; 0.000]
inclow		0.001
		[−0.006; 0.009]
Deviance	2571.839	2258.026
Dispersion	0.237	0.235
Num. obs.	10870	9620

E Robustness checks III: 2nd survey in 2016

E.1 Survey information

The survey was conducted in the French-, Italian- and German speaking parts of Switzerland between April and August 2016 (after a pre-test in February 2016) and implemented by the survey company LINK. It contains 1947 fully completed interviews. The sampling strategy differed considerably from the first survey: while recruitment was done via CATI based on the national telephone register for the first survey, we relied on the national official register for this second survey. Respondents were recruited via postal letter in which they were given a personalized login for completing the survey. Respondents were – if needed – reminded three times (via letter twice and a third time – if a phone number was available in the national official register – via telephone). Our sampling strategy was based on quota for the region, age and gender, drawn from the national census. Our overall response rate was 42%.

E.2 Specification of reform elements and levels of the conjoint design, 2nd survey in 2016

Table 7: Reform elements that are being discussed (values) (2nd survey, 2016)

Reform elements	**Levels**		**Goal of the reform elements**
Pension cutbacks 2nd pillar	1: status quo	No cuts (6.8% conversion rate)	**Retrenchment**
	2: government proposal	Cutbacks to 6%. Balance the lowering of pension levels by having people contribute more.	
	3: beyond gvt	Cutbacks to 6% No balancing.	
Cutbacks in widows' pensions	1: status quo	All widows below 64 are eligible for benefits	**Retrenchment**
	2: government proposal	Only widows with children <16 years should be eligible	
	3: beyond gvt	Stepwise abolishment of widows' pensions	
Increase in age of retirement	1: status quo	64 for women, 65 for men	**Retrenchment**
	2: government proposal	Increase for women by 1 year: 65 for both	
	3: beyond gvt	Stepwise increase for both men & women to 67	
Increase in the level of basic pensions	1: status quo	No increase in the level of basic pensions	**Compensation: targeting**
	2: government proposal	Increase by 70 CHF/month. In return: increase of contribution-payments by 0.3 percentage points	
	3: beyond gvt	Increase by 70 CHF/month	
Extended elibility 2nd pillar	1: status quo	No change. Only people earning >24'000 CHF/year are eligible	**Compensation: recalibration**
	2: government proposal	Extended access for part-time workers	
	3: beyond gvt	Extended access for people with lower incomes and part-time workers	
Increased revenues (VAT)	1: status quo	No increase in VAT	**Compensation: increased revenues**
	2: government proposal	Increase of VAT by max. 1 percentage points	
	3: beyond gvt	Increase of VAT by max. 2 percentage points	

E.3 Results conjoint analysis 2nd survey, 2016

Figure 4: Effects of reform elements on support for the pension reform package, pooled

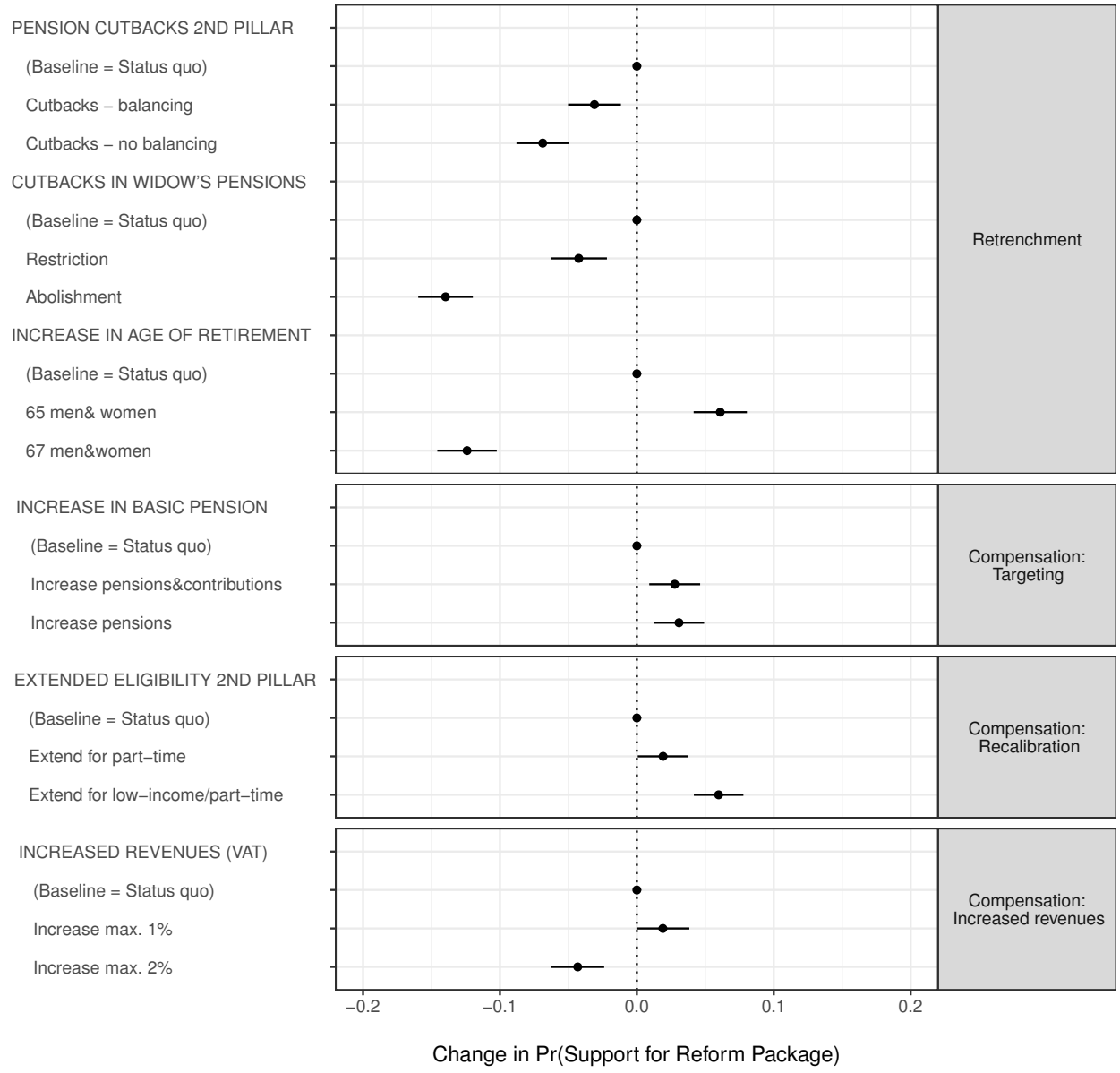


Figure 5: Effects of reform elements on support for the pension reform package, by income

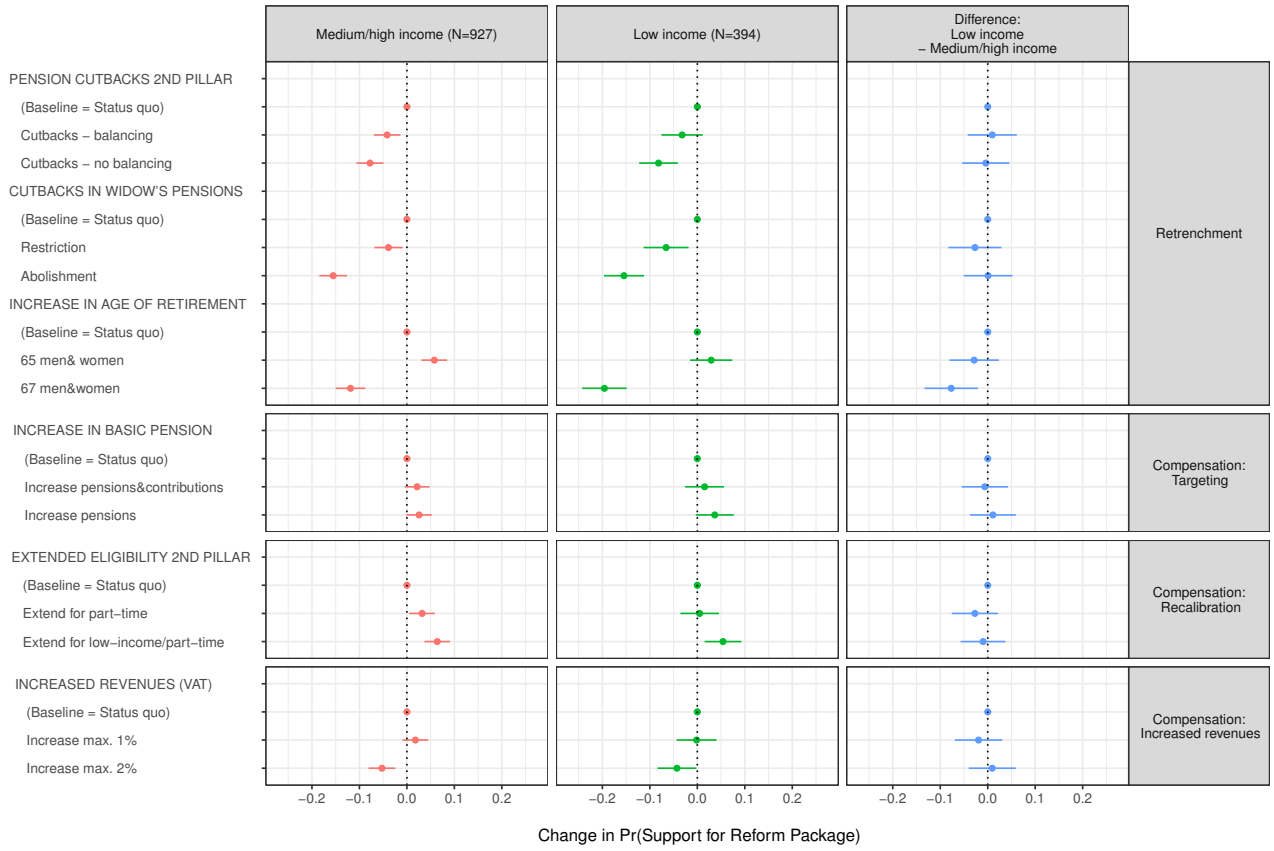


Figure 6: Effects of reform elements on support for the pension reform package, by gender

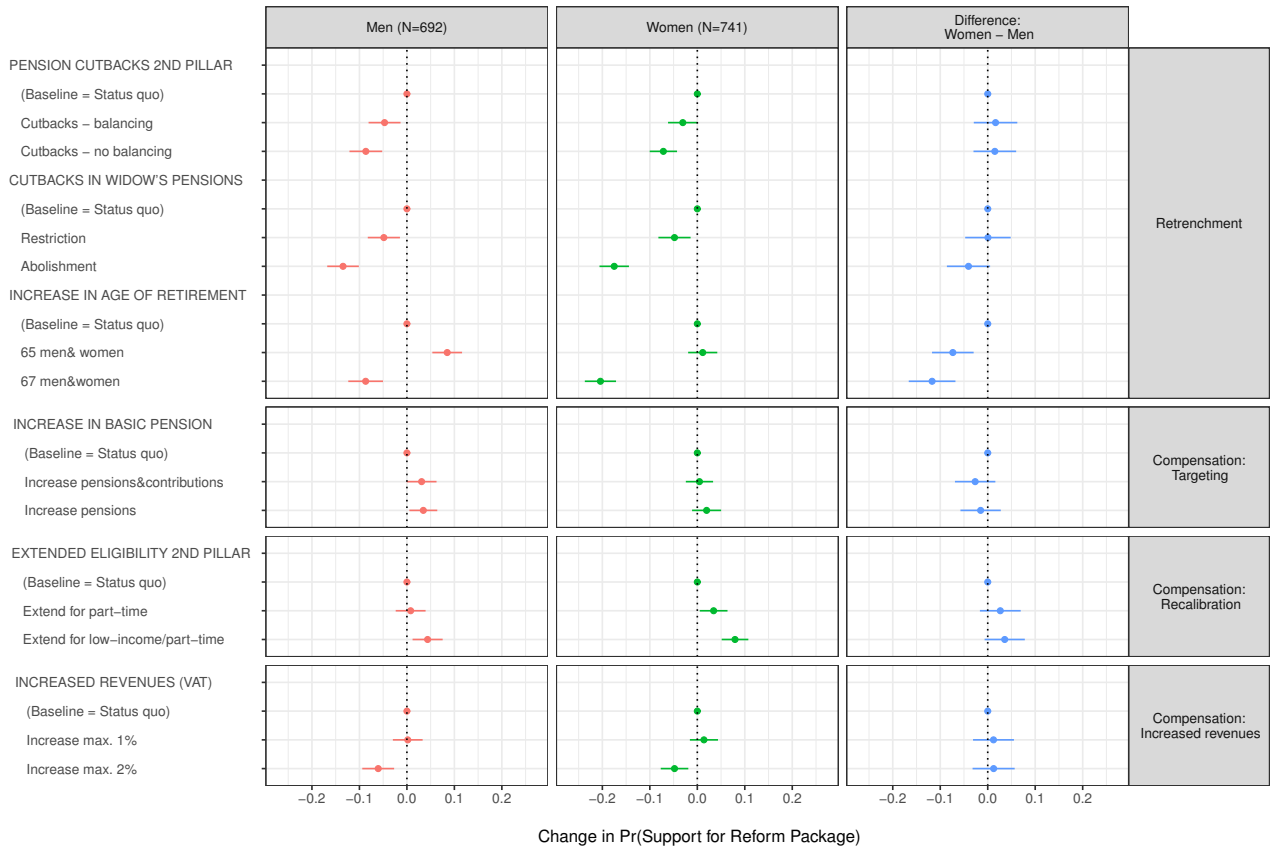
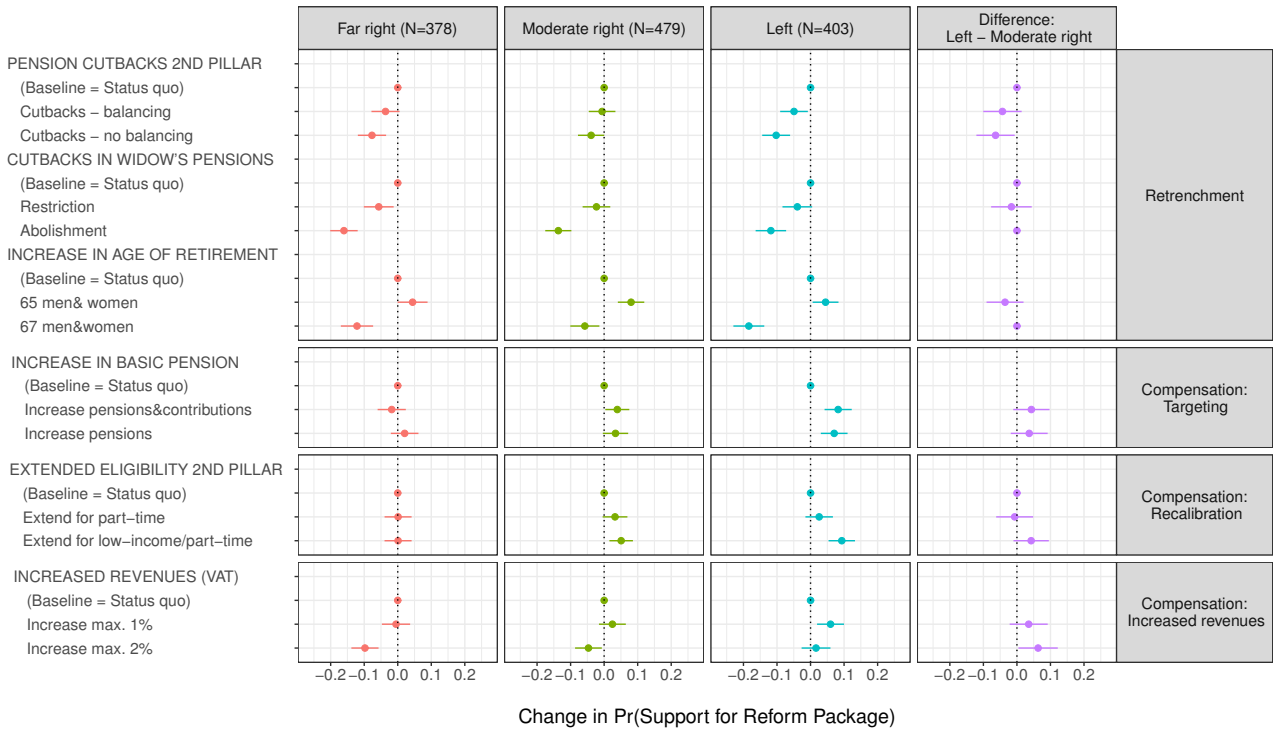


Figure 7: Effects of reform elements on support for the pension reform package, by party



F Robustness checks IV: Sophistication

To check whether the results differ between unsophisticated and sophisticated voters, we performed several tests. Sophistication was measured by tertiary education (= 1) vs. no tertiary education (sophistication = 0).

F.1 Split-sample test

Since the distribution of the sophistication variable is highly skewed (sophisticated = 622, unsophisticated = 1244), we draw repeated random samples and compare the average of computed correlation measures. More precisely, we use the following procedure:

Step A, unsophisticated voters:

1. Select random sample of 500 respondents
2. Select first split of packages (e.g. conjoint reform package comparisons 1-3)
3. Estimate AMCEs
4. Select second split of packages (e.g. conjoint reform package comparisons 4-5)
5. Estimate AMCEs
6. Calculate correlation between AMCE-estimates of split 1 and split 2 and save correlation
7. Repeat this procedure with 1000 random samples, each time saving correlation between AMCE estimates
8. Calculate mean of 1000 correlation measures

Step B, sophisticated voters:

Repeat 1-8, calculate mean correlation.

Step C, compare correlation between AMCEs:

Table 8: Correlations between AMCEs, sophisticated and unsophisticated respondents

Splits	Full sample	Sophisticated full sample N=622	Unsophisticated full sample N=1244	Sophisticated mean AMCE correlations of 1000 random samples of N=500	Unsophisticated mean AMCE correlations of 1000 random samples of N=500
(1,2) : (3,4,5)	0.96	0.89	0.96	0.87	0.89
(1,2,3) : (4,5)	0.95	0.82	0.97	0.80	0.91
(1,2) : (4,5)	0.94	0.85	0.96	0.83	0.88

F.2 Coherence between conjoint and direct questions

A very broad array of questions in our survey allows for an alternative way to examine potentially varying levels of comprehension among respondents. Beyond their choice of a reform package in the conjoint experiment, respondents were also asked about their attitudes towards several of the reform components in standard, uni-dimensional survey questions (Likert scale). For a total of three levels belonging to three different reform elements in the conjoint setting, we have sufficiently similar direct questions asked later in the survey (increase in retirement age, pension cutbacks second pillar, increase in VAT). We exploit this duplication to compare average within-respondent coherence between the group of sophisticated and unsophisticated respondents.

Step A: Individual attitudes in conjoint

Separate linear probability models are used to calculate respondent-specific estimates of the effect of the three specific values of interest (increase in retirement age to 67, pension cutbacks second pillar without compensation, increase in VAT by max. 3%) on choosing the displayed reform package or not. Given the small sample size per respondent (N=10), this obviously results in imprecise estimates. However, the point estimate nevertheless gives an indication regarding a respondent's stance towards the specific reform component. The resulting coefficients are subsequently classified into quartiles in order to a) match the coding of the direct question and b) avoid over-interpretation of imprecise estimates.

Step B: Individual attitudes in direct questions

The answer category to the direct questions asking about the same reform components ranges from 1 to 4 and is recoded to match the direction of the equivalent items in the conjoint setting.

Step C: Compare level of within-respondent coherence between groups

Two different measures are used to compare coherence levels between the sophisticated and unsophisticated group of respondents. First of all, Pearson's product moment correlation coefficient tests the association between the paired sample. An asymptotic confidence interval is given based on Fisher's Z transform. As an alternative, Krippendorff's alpha, a measure of coder reliability, is adapted to the purpose of comparing coherence between groups. The main interest is in the difference between groups. The absolute level of Krippendorff's alpha in this application is not particularly informative as the measure only evaluates whether ratings in both questions types (conjoint and direct) are identical (e.g. 4 and 4) and does not reward similarity (e.g. 3 and 4 as opposed to 1 and 4). Bootstrapping (2000 iterations) provides confidence intervals for the given probabilities.

The following table presents mean values and confidence intervals of both measures of comparison resulting from 1000 repeated random samples of each group (N=500) to avoid differences in the measures based on unequal group size.

Table 9: Coherence between conjoint and direct questions

Sample	Estimate	95% CI
Pearson's r		
full	0.463	0.422 - 0.502
sophisticated	0.482	0.442 - 0.520
unsophisticated	0.450	0.409 - 0.490
Krippendorff's alpha		
full	0.174	0.238 - 0.110
sophisticated	0.195	0.260 - 0.130
unsophisticated	0.161	0.226 - 0.097

Irrespective of the trusted indicator, as one would expect, coherence is slightly higher among respondents with tertiary education but the measures of coherence do not differ in statistically significant terms.