Public Opinion on Welfare State Recalibration in Times of Austerity: Evidence from Survey Experiments[*]

Björn Bremer[†]   Reto Bürgisser[‡]

December 2021

Forthcoming in Political Science Research and Methods

Abstract

Even though social investment is highly popular, welfare state recalibration remains an uphill battle. When resources are scarce in austere times, welfare recalibration involves multidimensional trade-offs. Existing research primarily studied preferences towards individual policies or trade-offs in specific policy fields, failing to capture citizens' overall social policy priorities. Using two novel survey experiments in three European countries, we show that citizens have clear social policy priorities: pensions and education enjoy a high, family policies a medium, and labor market policies a low priority. However, policy constituencies differ in their relative priorities. Our findings suggest that welfare state recalibration is difficult because trade-offs are unpopular, and distributive conflicts in mature welfare states are mainly about distributing resources to specific social groups.

Keywords: Welfare state; social investment; austerity; public opinion; survey experiments

---

[†]Max Planck Institute for the Study of Societies, Cologne. E-mail: `bremer@mpifg.de`

[‡]University of Zurich, Institute of Political Sciences. E-mail: `buergisser@ipz.uzh.ch`

1 Introduction

In the thirty years after the Second World War, one of the most remarkable features of politics in advanced economies was the constant expansion of the welfare state. The trend towards ever more generous and encompassing social protection came to a halt with the economic crises in the 1970s. As governments ushered in an "era of permanent austerity" (Pierson, 1996), the focus shifted from welfare state expansion towards cost containment and retrenchment. The concomitant structural change towards a postindustrial economy changed the economic context conditions and social policy. Rising new social risks revealed institutional frictions in a consumption-oriented welfare state built on the premise of an industrial society (Taylor-Gooby, 2004; Bonoli, 2007; Armingeon and Bonoli, 2007).

In order to cope with new social risks, governments have been forced to recalibrate social policy towards a more investment-oriented welfare state by incorporating new policy instruments, such as active labor market policy (ALMP), family policies, and educational policies (Morel et al., 2012; Hemerijck, 2013; Beramendi et al., 2015).[1] Extant research has demonstrated that these social investment policies are highly popular and allow for affordable credit-claiming (Ansell, 2010; Bonoli, 2013; Busemeyer and Garritzmann, 2017; Neimanns et al., 2018). However, despite their apparent popularity, welfare state recalibration often resembles a "political uphill battle" (Hemerijck, 2018), and many countries have not fully embarked on a social investment path (Morel et al., 2012; Bouget et al., 2015). Why is welfare state recalibration so difficult?

In times of limited resources, policymakers face tough choices about how to recalibrate the welfare state. Hard fiscal constraints combined with rising new social demands have transformed welfare state politics from a positive- to a zero-sum game (Levy, 1999; Häusermann, 2010). Governments struggle to increase spending on new social policies while maintaining traditional social protection systems. Consequently, welfare state reforms since the 1990s have often increased spending on some social policies at the cost of others (Bürgisser, 2019). As austerity became the dominant macroeconomic policy response to the Great Recession in post-crisis

---

[1]Social consumption policies hedge against traditional social risks such as unemployment, sickness, disability, or old age. Future-oriented social investment policies aim to prepare ex-ante to cope with structural change rather than repairing damage ex-post through compensation (Beramendi et al., 2015).

Europe, policymakers' room for fiscal maneuver has been even further reduced. This has put distributive conflicts center stage, as most government social spending is highly popular among the broader public (Häusermann et al., 2021).

We argue that welfare state recalibration under austerity is complicated because citizens, on average, do not strongly prioritize social investment over social consumption. Existing research on support for social investment mainly studies attitudes towards individual social policies or individual policy fields and does not capture the numerous trade-offs inherent in the multidimensional recalibration of the welfare state in times of austerity. This is unrealistic because the debate among policymakers and scholars alike is not anymore about more or less welfare state spending *per se* but the relative importance of different social policies. We thus shift the focus from citizens' social policy *positions* towards their social policy *priorities*, contributing to an emerging literature studying attitudes towards policy trade-offs (Busemeyer and Garritzmann, 2017; Neimanns et al., 2018; Häusermann et al., 2018; Armingeon and Bürgisser, 2021; Bremer and Bürgisser, 2021; Häusermann et al., 2021). Unlike previous research, we move beyond a few narrow social policy trade-offs and adopt a "macro-perspective": we ask what policy *priorities* citizens have regarding the *multidimensional* recalibration of the welfare state and study trade-offs within *and* across the most relevant social policy fields simultaneously.

Traditional, unidimensional survey questions only capture citizen's unconstrained policy *position* net of their importance. However, policy *priorities* result from both positions *and* importance, which are vitally important in times of tight budget constraints. Using novel data from two survey experiments conducted in Germany, Italy, and the United Kingdom in 2018, we study how citizens evaluate reform proposals when forced to choose between several highly popular social policies. First, we use split-sample experiments to tease out differences in citizens' policy positions and priorities towards different social investment policies (education, active labor market policies, and childcare). Second, we use a conjoint survey experiment that simultaneously varies spending on multiple social policy areas (pensions, education, passive and active labor market policies, childcare and child benefits) to determine the public's *multidimensional priorities* regarding welfare state recalibration. We employ a novel method to analyze results from this conjoint experiment to shed light on how citizens evaluate and prioritize different forms of welfare state recalibration in times of austerity.

The results reveal that citizens' policy positions and policy priorities are profoundly different. The split experiment shows that unconstrained support for social investment in three different policy areas is very high (policy position), but support drops substantially when citizens are confronted with constraining trade-offs (policy priorities). The conjoint experiment taps into multidimensional priorities and indicates that, apart from education, social investment is not a priority. More specifically, we demonstrate that narrow policy constituencies could play a crucial role in blocking welfare state recalibration. Specific social policies create their own support coalitions because citizens' narrow self-interest shapes their social policy priorities: people react strongly to cuts in spending from which they benefit. Our findings show which policy constituencies prefer which type of welfare recalibration and how strongly they resist welfare state retrenchment in specific social policy fields. For example, retired individuals prioritize pension spending more compared to other respondents, while people with children prioritize family policies more than others. In times of limited resources, these policy constituencies divide existing political coalitions, and, therefore, there is no clear social investment coalition that pushes for welfare state recalibration.

Our findings have several important implications. Substantively, they suggest that distributive conflicts in mature welfare states have become highly fragmented and are more about the distribution and protection of resources to specific groups than general support for the welfare state. Contributing to existing research (e.g. Busemeyer and Garritzmann, 2017; Häusermann et al., 2021), we show which groups most strongly resist welfare state recalibration, and which trade-offs are most contentious. Successful welfare state reforms need to be carefully crafted, including compensation to these policy constituencies that see their benefits at risk. Methodologically, our findings show that simple unidimensional preference questions exclusively capture an individual's policy position. They do not allow for inferences about respondents' policy priorities towards the multidimensional nature of welfare state recalibration. Our study uses two experimental approaches to capture such priorities. In particular, we propose a novel way to conduct and analyze conjoint experiments in order to shift the focus of public opinion research from *positions* towards *priorities*, and this approach could be applied to other policy fields to study salient trade-offs.

The remainder of this paper is structured as follows. First, we briefly review the existing literature and develop the theoretical expectations about the preferred forms of welfare state

recalibration. The second part describes the two survey experiments, and the third part presents the results. The final section concludes with a discussion of the broader implications of the findings.

2   Public Opinion on Welfare State Recalibration

2.1   From Unidimensional Positions to Multidimensional Priorities

Extant literature has shown that social policy preferences cluster along different dimensions, i.e., preferences towards social investment and social consumption are distinct (Fossati and Häusermann, 2014; Garritzmann et al., 2018). Much like support for the welfare state in general, social consumption and social investment are highly popular (Ansell, 2010; Roosma et al., 2013; Neimanns et al., 2018; Häusermann et al., 2021). This high support for social consumption and social investment is also corroborated by our survey results in Figure 1 (for more details on the survey, see the research design section). Citizens are strongly in favor of governments providing a safety net when people become ill, old, or unemployed (social consumption) but they also want governments to support people getting better education and new jobs (social investment). Neimanns et al. (2018) even show that social investment is far more popular than social consumption, arguing that social investment appeals to a broad coalition of well-educated middle class voters and is a valence issue.

However, despite the popularity of social investment, welfare state recalibration is politically difficult (Hemerijck, 2018). Austerity is a decisive factor why many countries have not recalibrated their welfare state towards social investment, particularly in Southern Europe (Bouget et al., 2015). Limited resources do not allow for across-the-board increases in welfare state spending and force governments to make tough choices. Due to political and economic reasons, they often cannot increase taxes or government debt to finance welfare state expansion on the scale common in the post-war period. On the one hand, government debt is seen as an illegitimate way to finance social policy spending, especially since the European sovereign debt crisis (Blyth, 2013; Barnes and Hicks, 2018). On the other hand, governments' ability to tax effectively has eroded over the past decades, either due to ineffective tax authorities, high capital mobility, or sophisticated tax evasion schemes (Zucman, 2015).
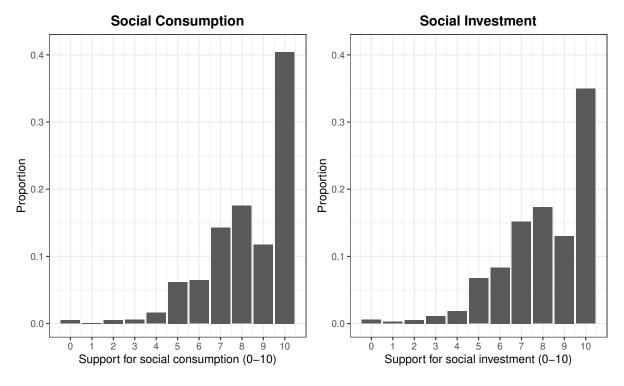
Figure 1: Support for social investment and social consumption

Note: Level of (dis-)agreement to the statement that "the government should provide a safety net when people become ill, old, or unemployed" (left panel) and that "the government should support people to get better education and new jobs" (right panel).

As it is almost impossible to increase social spending on all policies in times of permanent austerity (Pierson, 1996), research has begun to study public opinion trade-offs. Boeri et al. (2001) demonstrated that support for unemployment benefits and pension spending is lower once citizens learn how much they have to pay for it through higher contributions or taxes. Likewise, more recent studies indicated that support for education and family policy spending does not reach a majority when citizens are confronted with the necessity to cut pensions or unemployment benefits in exchange (Busemeyer and Garritzmann, 2017; Neimanns et al., 2018). Finally, other research has studied multidimensional trade-offs within specific social policy fields, such as pensions or labor market policy (Gallego and Marx, 2017; Häusermann et al., 2018). While these studies have significantly advanced our knowledge on the relevance of policy trade-offs, welfare state recalibration is a more complex and broader endeavor. We still lack coherent information on two-dimensional trade-offs across the whole range of social policy trade-offs and how they influence support levels. In addition, research studying multidimensional trade-offs has exclusively focused on single policy fields. They only allow us to gauge the relative popularity of one or two social policy areas and fall short of capturing the true *multidimensional* nature of welfare state recalibration. Ample research has demonstrated that welfare state reforms

increasingly combine multiple policy areas simultaneously (Seeleib-Kaiser, 2008; Häusermann, 2012; Hemerijck, 2017). Therefore, we adopt a macro-perspective and study how citizens evaluate six different social policies against each other and weigh the trade-offs associated with welfare state recalibration across social policy fields.

Welfare State Recalibration and the Public's Social Policy Priorities

In this paper, we focus on trade-offs between the most relevant social policies in a context of austerity: pensions, education, active and passive labor market policies, and childcare and child benefits. We argue that policymakers have to balance trade-offs across multiple social policy areas and craft reform packages that prioritize some policies over others. Are citizens ready to reduce spending on pensions to increase spending on education and childcare? Are citizens willing to pay for childcare by lowering child benefits? Do citizens prioritize labor market policies over childcare?

In principle, support for social investment is very high among the wider public. However, we expect that this broad support substantially declines when people have to evaluate trade-offs associated with welfare state recalibration. Trade-offs often occur between equally cherished policies, forcing people to weigh the benefits and costs in question carefully. The way they make this evaluation should differ across individuals (see below), but we expect the average support for social investment to drop when trade-offs are mentioned. The diffuse and broad support coalition for social investment (Neimanns et al., 2018) might particularly easily fall apart in light of social policy trade-offs with clearly defined beneficiaries.

Hypothesis 1. *Support for social investment declines and is not a priority when it entails trade-offs with other social policies.*

Although we expect average support to decline drastically for all social policy fields once we introduce trade-offs, there should be differences across social policies. First, the average citizen should be more supportive of increasing spending in fields that benefit many people because they are more likely to profit from it. Conversely, the average citizen should react most sensitively to retrenchment in those fields. This involves spending on both social consumption and social investment. For example, pension spending is likely the most popular form of social spending

6

because most people expect to retire and to receive a pension at some point. Similarly, education is a central pillar of the welfare state that directly affects many people. Labor market policies and family policies, in contrast, benefit a smaller group of recipients.[2]

Apart from sheer group size, a second relevant dimension is "deservingness" perceptions. The literature shows that particular social groups are seen as more or less "deserving" of welfare state support (van Oorschot, 2006). For example, elderly people are considered most deserving because retiring is common, and they have earned their right to be supported by having worked and contributed to society. In stark contrast, the unemployed are seen as much less deserving. There is a significant stigma associated with unemployment, which is often seen as a result of one's passivity or unwillingness to find a job and a failure to contribute to society (Jensen and Petersen, 2017).

Although deservingness perceptions may reflect the size of a particular welfare state constituency, they also tap into normative considerations, which help distinguishing support for groups that are equal in size. There is little literature on deservingness perceptions of families, but we assume that they are seen as more deserving than the unemployed. In contrast to retirement, having children is still a choice, albeit less stigmatized than being unemployed. Therefore, we expect that citizens have social policy priorities based on the size and the deservingness of the beneficiary group.

Hypothesis 2. *Pension and education have a higher priority than family policies, which, in turn, have a higher priority than labor market policies.*

These priorities should influence the extent to which support for social investment drops in two-dimensional trade-offs, but they should also influence support for multidimensional welfare state reforms. Areas that citizens prioritize involve social consumption (e.g., pension) and social investment policies (e.g., education), reducing overall support for recalibration towards a social investment welfare state.

---

[2]In our survey, 21 percent of respondents are pensioners while 45 percent benefit from education (being in education or having children). As many will retire at some point, the support group for pension is likely much higher. The constituency for labor market policies (outsiders) and family policies (people with young children) are 23 and 21 percent, respectively.

Heterogeneous Priorities: The Importance of Policy Constituencies

The way that individuals respond to trade-offs should also vary across individuals. Citizens do not only evaluate policies based on their expected aggregate effects, but they also consider their pocketbook consequences. An extensive literature on unidimensional policy preferences has highlighted important predictors such as income, class, education, socioeconomic status, and political ideology. In particular, the literature emphasizes material self-interest (income and risks) as a primary source of preference formation (Meltzer and Richard, 1981; Rehm, 2011). However, trade-offs involved in welfare state recalibration often occur between two equally cherished policies. Broad material self-interest may, therefore, be less relevant to explain social policy priorities in trade-off settings.

We expect that distributive conflicts in mature welfare states have become more about distributing resources to narrow groups than about support for the welfare state in general (see also, Busemeyer and Neimanns, 2017; Busemeyer and Garritzmann, 2017). Welfare states are multidimensional entities from which individuals benefit in various ways. Building on the policy feedback literature (e.g., Pierson, 1993), we expect that specific social policies create their support coalitions, which we call policy constituencies. While policy constituencies may generally not have a higher *position* for policies they directly benefit from, they should have a higher *priority* for "their" policies. When fiscal constraints come to bite, trade-offs between different social policies are exacerbated. In this context, citizens should be primarily interested in protecting the services and benefits they directly benefit from.

Overall, specific policy constituencies should react more strongly to trade-offs that affect them (more or less) directly. Citizens thus are likely to prioritize those policies where they expect to be net beneficiaries (see also Rehm, 2011). In some cases, these policy constituencies are neatly defined. Having young children should make individuals more sensitive to trade-offs that involve a reduction of family policies. Similarly, labor market outsiders should have a keen interest in protecting unemployment benefits and training for the unemployed that help improve their situation directly (see also, Rueda, 2007; Schwander and Häusermann, 2013; Vlandas, 2020). Finally, retired people should be more sensitive to changes in pension spending and less interested in trade-offs that involve policies that do not affect them anymore (e.g., family policy). In other instances, policy constituencies are fuzzy. Education benefits a large and diverse group

of people, including university students, young adolescents in school or vocational training, and parents with children. Moreover, education is an investment in human capital. Its benefits are often uncertain and mainly accrue in the future. This should create a diffuse support coalition for education with only small constituency-specific effects.

Hypothesis 3. *Policy constituencies prioritize policies from which they benefit more than people not benefiting from these policies.*

Research Design

We fielded a survey in three large European countries in 2018: Germany, Italy, and the United Kingdom (UK) to test our expectations. The countries represent major European economies with advanced welfare states, including different varieties of capitalism (Hall and Soskice, 2001) and different welfare state regimes (Esping-Andersen, 1990). They spend different resources on different social policies, which gives us sufficient variation to check whether individual countries drive our results. In each country, we recruited 1,200 respondents from a pool of eligible voters through large online panels provided by Qualtrics. The sample was representative of all eligible voters based on gender and age.[3]

To assess citizens' *multidimensional social policy priorities*, we used two different survey experiments that we developed to overcome problems with conventional surveys studying one- or two-dimensional trade-offs. First, we use a split-sample experiment to measure individuals' support for social investment given different kinds of trade-offs, allowing us to zoom into trade-offs within and across policy fields to gauge support for welfare state recalibration. Second, we use a conjoint survey experiment that simultaneously varies six forms of social spending to measure the public's macro priorities across the full range of social policies. Both experiments impose austerity to avoid "free lunches" where a respondent can support spending increases on all social policies. This permits us to assess which social policies they prioritize over others. We only focus on social policy priorities because increasing revenues through taxes or government debt are often difficult. Moreover, we want to study attitudes toward direct recalibration, i.e.,

---

[3]Before fielding the survey, the research design was pre-registered and received ethical approval.

shifting resources from one social policy field to another, and not indirect recalibration through welfare state expansion financed by tax or debt increases.

Split-Sample Experiment

To test how respondents perceive and react to salient trade-offs associated with welfare state recalibration, we confronted respondents with a series of questions that measure individuals' support for three social investment policies. Respondents were randomly assigned to four different groups, including one control group and three treatment groups.[4]

In each group, respondents evaluated three statements about spending on education, active labor market policies, and childcare. For each statement, respondents in the control group were presented with traditional, one-dimensional questions, while the treatment groups were confronted with statements that raised awareness for different two-dimensional trade-offs. Subsequently, they were asked to what extent they agree or disagree with these statements. This allows us to analyze the extent to which specific social policy trade-offs are more acceptable to respondents than others.

Table 1 shows the complete list of statements. All statements are relative to the status quo and do not quantify the specific trade-offs. This simplification allows us to run the same experiment in several countries. To reduce complexity and cognitive exhaustion for the respondents, we limited the trade-offs to those theoretically most interesting. We always included an intra-field trade-off (e.g., ALMP vs. PLMP, childcare vs. child benefits). Additionally, we always introduced a trade-off between pension and childcare spending because these two are among the most popular forms of consumption and investment spending. Finally, in the case of education, we also introduce a trade-off with ALMP.

---

[4]The conjoint survey experiment occurred before the split-sample experiment in the survey to avoid possible experimental bias across the control and treatment groups.

Table 1: Design of the split sample experiment

Split 1 (Control)	Split 2 (Treatment 1)	Split 3 (Treatment 2)	Split 4 (Treatment 3)
The government should increase spending on education.	... even if that implies lower spending on old-age pensions.	... even if that implies lower spending on training schemes for the unemployed.	... even if that implies lower spending on childcare services.
The government should increase spending on training schemes for the unemployed.	... even if that implies lower spending on unemployment benefits.	... even if that implies lower spending on old-age pensions.	... even if that implies lower spending on childcare services.
The government should increase spending on childcare services.	... even if that implies lower spending on child benefits.	... even if that implies lower spending on old-age pensions.	... even if that implies lower spending on unemployment benefits.

2.2 Conjoint Experiment

Welfare state recalibration often involves multidimensional trade-offs across several policy fields simultaneously. To study priorities in a multidimensional setting, we designed a conjoint survey experiment that focuses on social policy trade-offs across several social policy fields. Conjoint survey experiments are useful to elicit priorities because they require respondents to evaluate entire policy packages rather than individual measures (Hainmueller et al., 2014). In our case, we evaluate the priorities that individuals attach to different social policies by forcing respondents to simultaneously make decisions based on the position and the intensity of their position towards multiple policies. Instead of asking respondents directly, we use the conjoint experiment to indirectly force respondents to reveal their social policy priorities. This allows us to capture priorities in one analysis rather than triangulating between results from different split-sample experiments jointly, as done above.

Specifically, our survey asked respondents to evaluate various reform proposals regarding changes to government spending in a set of choice tasks. Each task presented respondents with two profiles of possible welfare state reforms, and they were asked five times (i) to choose between two fiscal packages ("choice" variable) and to (ii) indicate how likely they are to support each of the packages ("rating" variable). The profiles comprised six attributes corresponding to

Table 2: Attributes and Levels of the Conjoint Experiment

	Attribute	Attribute Levels
Investment	Training for the unemployed	Decrease spending
		No change
		Increase spending
	Childcare services	Decrease spending
		No change
		Increase spending
	Education	Decrease spending
		No change
		Increase spending
Consumption	Unemployment benefits	Decrease spending
		No change
		Increase spending
	Child benefits	Decrease spending
		No change
		Increase spending
	Old-age pensions	Decrease spending
		No change
		Increase spending

core elements of welfare state spending, and each attribute could take on a set of discrete and pre-defined levels, representing different spending options. The profiles were then generated randomly, i.e., they contained a fixed number of attributes, which were shown to respondents in random order and with a random display of an attribute level.[5]

To reduce complexity and avoid cognitive exhaustion, we had to limit the number of attributes and levels. Concretely, the reform profiles contained six attributes (as shown in Table 2), chosen to represent two different dimensions of government social spending, namely social investment and social consumption. Education, childcare services, and training for the unemployed (ALMP) are proxies for social investment spending. Pensions, child benefits, and unemployment benefits (PLMP) are proxies for social consumption spending. The profiles did not include health care because of vast country differences in healthcare funding, making comparisons difficult, and spending on health care has both an investment and consumption element.

We included three levels for each spending attribute (increase, decrease, no change). In a fully randomized setting, there would be 729 combinations. However, to focus on trade-offs and priorities and avoid "free lunches" where respondents can favor spending increases on all

---

[5]The full instructions for the conjoint tasks are in Appendix C. The attribute order was randomized but held constant within individual respondents across the five tasks.

social policies, we force respondents to decide which social policies they prioritize over others. We achieve this by showing respondents only zero-sum reform proposals, i.e., we only allow combinations in which any given expenditure increase in one attribute is matched with an expenditure decrease in another attribute. This imposes austerity: it does not give respondents the option to choose expansionary reforms and thus makes the trade-offs inherent in welfare state recalibration under fiscal constraints explicit.

Respondents may disagree with these restrictions. Some respondents may want to increase spending on all social policies, while others may want to decrease spending on all of them. However, the zero-sum reform proposals allow us to tease out respondents' policy priorities on welfare state recalibration in times of austerity, when a (more or less) budget-neutral reform is politically most likely.[6] Consequently, we exclude 588 combinations and are left with 141 possible combinations. Importantly, the likelihood that a given level appears in combination with another level is still the same as if there were no restrictions because logical inconsistencies were uniformly deleted. This crucial feature of our design stems from the fact that each attribute includes three symmetrical levels across all attributes (increase, decrease, no change). Each package is randomly created within this subset of consistent packages and has the same likelihood to occur.

We calculate two main variables of interest from the conjoint experiment. First, we identify the causal effect of individual attribute levels on the support for the entire policy package compared to the baseline attribute level. To this end, we estimate the average marginal component effect (AMCE) of a change in the value of one of the six dimensions on the probability that the respondent chooses the reform proposal. This "choice" variable is binary: 1 if a reform proposal is chosen and 0 if not chosen. We estimate the AMCEs using linear probability models and regress the dependent variable on dummy variables for level (where the status quo is the baseline for each dummy). The ACME's incorporates both the position and the importance that individuals assign to each attribute level (Bansak et al., 2020) and, therefore, captures what we conceptually understand as policy priorities. Second, we follow Leeper et al. (2020) and measure

---

[6]Theoretically, our strong balanced budget assumption could be relaxed due to the possibility of a Keynesian multiplier. However, the time lag associated with government spending (and the lag on its returns, in particular) makes it still necessary for governments to scale down specific policies to spend more on others, especially in times where additional revenues through tax or debt increases are very unlikely.

subgroup differences by calculating recommendations by calculating conditional marginal means for all levels of our attributes. Marginal means measure how favorable respondents are to a given feature of our reform package, which allows us to analyze differences across policy and party constituencies.

To estimate the AMCEs and marginal means, we use Ridge regression. Standard conjoint experiments have dimensions that are independent and fully randomized. Spending trade-offs are not independent by design: under austerity, increasing expenditure on one policy requires decreasing expenditure on another policy. Therefore, each attribute's value depends on the other attributes to ensure to ensure a fully balanced reform. To account for these dependencies, we use a novel approach that uses ridge regression. Ridge regression is a standard regularization method that Hoerl and Kennard (1970) proposed to address design-based super collinearity and that Horiuchi et al. (2018) also used for conjoint analysis. To estimate the AMCEs, we use the R package `glmnet`, and we rely on bootstrapping to calculate non-parametric confidence intervals. This approach allows us to make inferences about the effect of an attribute changing from one value to another, averaging over the randomization distribution of our 141 profiles. We explain the method and rationale further in Appendix D. Moreover, we discuss several robustness tests at the end of the results section, which did not change the results.

3   Results

3.1   Two-Dimensional Priorities of Welfare State Recalibration

To study how respondents react to specific trade-offs, we use results from the split-sample experiment. We first graphically present the predicted mean support (on a scale from 0 to 10) and 95 percent confidence intervals based on OLS regressions. Figure 2 thus shows the predicted support for spending increases in three social policy fields without a trade-off (control group) and with three different trade-offs (treatment groups). To test the robustness of our findings, we control for several covariates (e.g., age, sex, occupational classes, income, education, employment status, partisanship, parents with children) and include country-fixed effects.

Figure 2: Average support for spending increases by treatment and model, pooled
Note: Predicted mean support and 95 percent confidence intervals based on OLS regressions. The dependent variable is scaled from 0 to 10. Covariates are age, gender, education, occupation, work contract, partisanship, and having young children.

In line with Figure 1, the results show high levels of support for social investment policies in an unconstrained setting. However, it is implausible that governments do not face any fiscal constraints and can increase spending on all social policies. Therefore, we forced respondents to make a distributional choice between similarly cherished social policies. Overall, and in line with existing research on trade-offs (Busemeyer and Garritzmann, 2017; Neimanns et al., 2018; Häusermann et al., 2021), our results indicate that the widespread support for social investment tumbles once we introduce trade-offs. These treatment effects are large and significant for all three policy fields. Moreover, the results are robust to the inclusion of covariates and country-fixed effects.

To estimate the share of respondents who support spending increases, we transform the 0-10 response scale into a binary variable. Responses from six to ten are counted as agreement, while responses from zero to five are counted as disagreement/neutral. Education is the most popular social investment policy. The average level of support for education is about 8.0, implying that around 88 percent of the respondents agree that the government should increase spending on
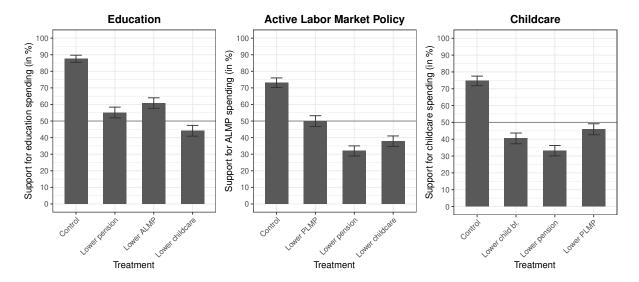
Figure 3: Share of supporters for spending increases by treatment
Note: Predicted share of respondents supporting spending increases and 95 percent confidence intervals. The dependent variable (0-10) is transformed into a binary variable (6 to 10 are counted as support; 0 to 5 as opposition/neutral.

education. ALMP and childcare enjoy slightly lower but still widespread support with mean support levels of 6.9 (73 percent) and 7.1 (75 percent), respectively.

The share of supporters for all social policies drops substantially when trade-offs are introduced, but Figure 3 shows some interesting variations across social policies. First, higher education spending remains relatively popular even when trade-offs are acknowledged. A majority of 61 percent supports education spending if it implies cuts to ALMP. Cuts to childcare are the least popular, with only 44 percent supporting higher education spending. Remarkably, still 55 percent support an expansion it if it implies pension cuts. This contrasts with results from (Busemeyer and Garritzmann, 2017), who show that an expansion of education at the cost of pension spending finds no majority support.

Second, ALMP spending has a lower priority than education spending, given that support falls more when trade-offs are introduced. However, almost a majority of respondents support higher ALMP spending at the cost of lower PLMP spending, indicating the possibility of re-calibrating labor market policy towards activation. Support for ALMP is lower and does not generate a majority when it implies pension cuts (32 percent), the most significant area of social consumption spending, or lower childcare spending (38 percent).

Finally, welfare state recalibration towards childcare does not find a majority, either. The most significant drop in the share of supporters occurs when it implies lower pension spending,

indicating that the average citizen is more concerned about protecting pension benefits than helping young families. Contrary to labor market policies, there is no support for an "activation turn" within family policies. Only a minority of respondents (40 percent) are willing to increase childcare spending at the cost of lower child benefits. Interestingly, only 46 percent of respondents support an expansion of childcare when this comes at the costs of unemployment benefits, which is usually a relatively unpopular form of spending.

Overall, the results show that welfare state recalibration is an uphill battle in times of limited resources. In principle, support for education, ALMP, and childcare spending is high, but they are not clear policy *priorities* in a two-dimensional trade-off setting. Politically, increases in education spending are the most viable. Moreover, a recalibration within some policy fields (from PLMP to ALMP) could be possible but likely contentious. A more pronounced social investment turn by spending fewer resources on pensions is certainly the least preferred and politically most difficult avenue of welfare state recalibration.

3.2   Multidimensional Priorities of Welfare State Recalibration

The results from the split experiment suggest that welfare state recalibration towards social investment is not a priority. However, welfare state reforms often impact several policy fields at the same time. We thus turn towards the conjoint experiment to account for this multidimensionality. Figure 4 shows the AMCEs of increasing or decreasing spending items relative to the baseline ("no change") on the probability (expressed in percentage points) that a given reform package is supported. The results indicate that the average respondent has clear priorities when confronted with trade-offs across all social policy fields.

First, spending on education and pensions is highly popular among the average citizen, even in the context of austerity. Additional pension spending increases the probability that respondents support the overall package by almost five percentage points compared to the status quo, while lower pension spending decreases the overall probability of support by more than seven percentage points. Education spending is only slightly less important to respondents. Lower education spending reduces support by almost four percentage points, whereas additional spending on education increases the support by a similar amount.
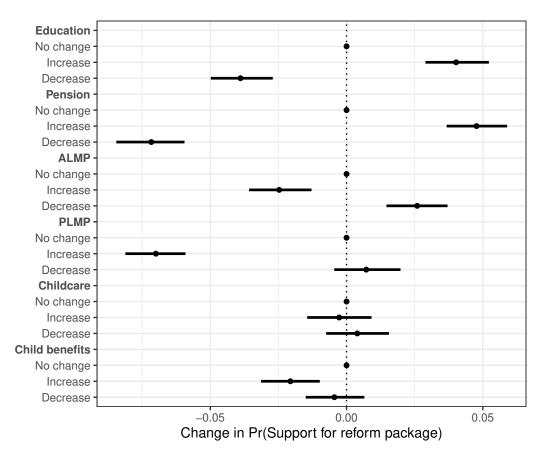
17

Figure 4: AMCEs from the conjoint survey experiment
Note: Average component-specific marginal effect (ACME) of a change in the value of one dimension on the probability to choose the reform package ("choice" variable).

Second, spending on labor market policies is not a priority in the context of austerity and sharp trade-offs. Given that expansionary reforms are impossible, respondents do not favor increasing spending on labor market policies, which would necessitate cuts in other areas. In particular, citizens react negatively to increasing spending on PLMP, which reduces support by around seven percentage points. ALMP spending is less unpopular, but it still reduces the support for a given reform proposal by 2.5 percentage points. Interestingly, the pattern reverses when it comes to decreasing spending on labor market policies. Lower ALMP spending increases support, whereas lower PLMP spending does not affect support. In sum, spending on the unemployed has a low priority for the average citizen. The average respondent is less supportive of reducing spending on ALMP than PLMP, but they certainly do not want to increase ALMP spending, either. Thus, the unemployed are likely viewed as less deserving than other social groups that benefit from the welfare state.

Third, family policies have a relatively small influence on the probability that respondents support a given package. On average, respondents oppose increasing spending increases on child benefits, while decreasing these benefits does not affect overall support. Surprisingly, neither an increase nor decrease in childcare affects overall support, indicating that they do not matter for evaluating policy packages. The average respondent seems ambivalent about family policies, except that increasing child benefits negatively influences reform support.

Overall, the results are consistent with the findings from the split-sample experiment. They suggest that citizens have clear social policy priorities in the context of austerity. First, they care most about broad policies that many respondents benefit from and thus attach a high priority to pension and education spending. Second, labor market policies enjoy only a low priority among citizens, as evidenced by the adverse effects of higher spending on such policies. The average citizen is especially opposed to higher PLMP spending. Third, the small to non-existent effects of family policies indicate that citizens attach a medium priority to family policies. They are more popular than labor market policies but less popular than pensions and education.

Relative Priorities Across Policy Constituencies

Welfare state politics is often influenced by small groups who benefit from specific policies. As theorized above, there are reasons to believe that the support for different social policies varies across policy constituencies due to narrow self-interest. To operationalize these policy constituencies, we identify the people most likely to benefit from specific social policies. We define the retired as the main policy-constituency for pensions, students and people with children for education, the labor market outsiders for both types of labor market policy, and people with young children for both types of family policy. We infer these characteristics from a battery of socio-demographic questions asked in the survey, and the operationalization of all independent variables is detailed in Appendix A.

Following our theoretical expectations, we calculate the marginal means for the policy constituencies compared to everyone else. Table 3 shows the marginal means for each dimension for the constituency and the non-constituency, the difference between the two, and whether this difference is statistically significant. The results show that specific policy constituencies react more sensitively to changes that involve specific policies from which they benefit directly. They

Table 3: Estimated marginal means of policy constituencies vs. non-constituencies

		Constituency	Non-Constituency	Difference
Education	No change	0.502	0.496	+0.005
	Increase spending	0.548	0.538	+0.010
	Decrease spending	0.449	0.466	-0.016*
Pension	No change	0.519	0.505	+0.013
	Increase spending	0.639	0.544	+0.094*
	Decrease spending	0.345	0.451	-0.105*
ALMP	No change	0.510	0.496	+0.014*
	Increase spending	0.464	0.466	-0.002
	Decrease spending	0.524	0.540	-0.016*
PLMP	No change	0.518	0.519	-0.002
	Increase spending	0.466	0.428	+0.037*
	Decrease spending	0.514	0.549	-0.035*
Childcare	No change	0.489	0.504	-0.016*
	Increase spending	0.515	0.484	+0.031*
	Decrease spending	0.498	0.511	-0.013
Child benefits	No change	0.501	0.510	−0.009
	Increase spending	0.519	0.470	+0.049*
	Decrease spending	0.480	0.519	-0.039*

Note: * p<0.05; constituencies are the retired for pension, the people in education and with children for education, the outsiders for PLMP and ALMP, and having young children for child benefits and childcare.

are more likely to support additional social spending than the average citizen when existing policies benefit them directly, and they are more likely to withdraw support if the reform involves cuts in "their" policies.

First, the marginal mean of increasing pension spending is 0.094 higher for retired respondents than non-retired. In other words, retired people have a 9.4 percentage point higher probability of supporting the reform if it includes an increase in pension spending than people who are not retired. Similarly, when a reform decreases pension spending, retired people have a 10.5 percentage point lower probability of support than non-retired people. These differences are statistically significant and substantial. Retired people will strongly oppose any cuts in pension spending. Going beyond the table, Figure 4 shows the marginal means for a selection of other conjoint attributes. Due to their higher support for pensions, it indicates that the retired are more likely to prioritize cuts in unemployment benefits than non-retired people when forced to make a choice.

Second, the constituency effects for education are smaller than for pensions. For education, the policy constituency has a one percentage point higher probability of supporting the reform if it includes an increase in education spending than the non-constituency. In comparison, it has a 1.6 percentage point lower probability of supporting it if it includes a decrease. Since the constituency benefiting from education is relatively diffuse and people who previously benefited from education (e.g., people with a university degree) likely remain supportive of it, it is not surprising that the heterogeneous effects are smaller than for pension spending. Interestingly, however, the policy constituency for education is much less likely to support pension spending increases.

Third, there are substantial differences between labor market outsiders and the rest of the population if the reform proposal increases or decreases PLMP spending. The non-constituency is firmly against an increase, whereas outsiders are much less opposed. Similarly, the non-constituency is much more favorable towards reforms that reduce unemployment benefits than outsiders. We find similar effects for decreasing ALMP spending, but there is no difference in support for increasing ALMP spending. Thus, outsiders are not more favorable towards activation reforms, because they associate them with strict conditionality and workfare policies (Fossati, 2018) or because a diffuse cross-class coalition of employers and the middle classes equally supports activation (Bonoli, 2013; Schwander, 2020). Outsiders are also much less supportive of increases in pension spending than insiders. They are less likely to receive a secure pension and are thus less likely to prioritize such spending.

Finally, people with young children are more supportive of additional spending on childcare and child benefits than all the other respondents. The marginal mean is 0.049 and 0.031, respectively, higher for people with young children than those without. This implies that they have a 4.9 and 3.1 percentage point higher probability of supporting childcare and child benefits, respectively. In contrast, when the reform package includes cutbacks in child benefits or childcare, they are less favorable towards such reforms than people without children. Much like the other narrow constituencies, people with children are thus concerned for their narrow self-interest when evaluating reforms in times of austerity. They, too, are less supportive of pension spending, but like all other policy constituencies, they are still unwilling to support cuts in pension spending.
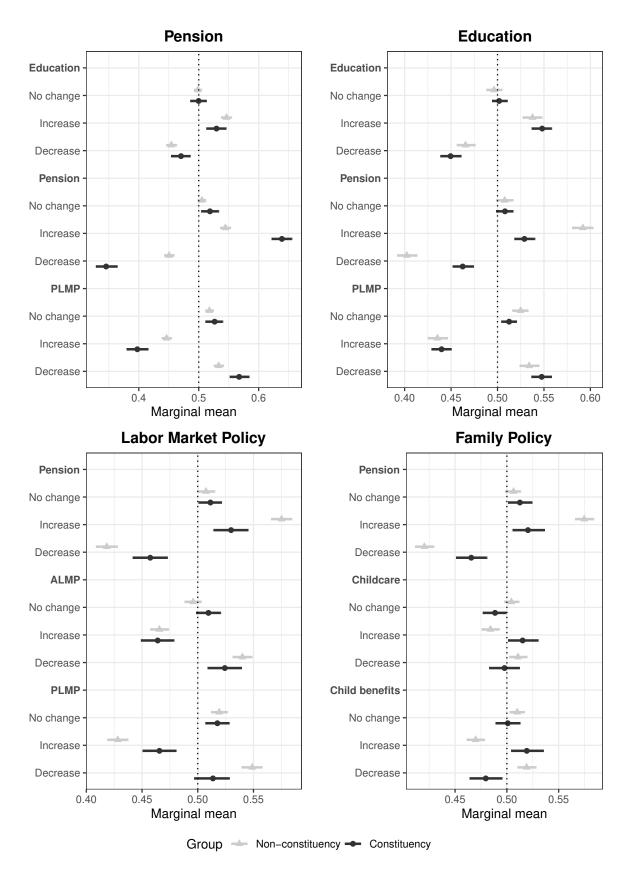
Figure 5: Estimated marginal means of policy constituencies vs. non-constituencies
Note: Estimated marginal means for the policy constituency vs. non-constituency (conjoint "choice" variable). For each graph, only a selected number of dimensions are shown. The complete graphs are included in Appendix G.

Overall, our findings suggest that narrow self-interest strongly influences support for welfare state recalibration.[7] Policy constituencies strongly support the policies they benefit from. Even policy constituencies that benefit from education spending, ALMP, and childcare are still not more likely to support a reduction in social consumption. In particular, the reluctance to reduce pensions is prevalent across all constituencies, making welfare state recalibration towards social investment in times of austerity unlikely.

Robustness Checks

Some concerns could be levied against our survey experiments. First, to ensure the validity of our sample, we included attention check and speeding checks, which automatically screened out respondents who paid no attention or sped through the survey (see Appendix B). Second, we used quota sampling based on age and sex, but we also checked that our sample is relatively representative of income and partisanship. Third, we matched the population's demographic characteristics in each country as closely as possible using entropy balancing (Hainmueller, 2012), and the results remain substantively the same (Appendix F).

So far, we have exclusively shown pooled results from three different countries. In Appendix E, we show that the general results hold across the three countries. Support for social investment spending declines in all three countries when it entails trade-offs, and priorities are similar in all three countries. Still, there are some minor differences. First, decreasing education spending is less popular in the UK than in the other two countries, while increasing education spending is most popular in Germany. Second, pension spending is relatively more important to respondents in Italy than in the other two countries. In turn, Italian respondents are much less opposed to increasing PLMP spending than respondents in Germany and the UK, while Germans are considerably more opposed to increasing ALMP spending than respondents in the other two countries. Finally, the UK is the only country strongly opposed to increasing child benefits, while family policy has no effect in Italy and Germany.

---

[7]Appendix G presents additional subgroup analyses, showing that differences between policy constituencies are larger than differences between income groups, educational groups, or party supporters.

We also used a set of tests to confirm that the results from the conjoint experiment are robust. First, we conducted a pretest with an opt-in sample from *Prolific* to analyze whether respondents were cognitively able to process a more complex conjoint task (eight attributes and more specific attribute levels). The results suggested that the design was slightly too complicated and that respondents preferred a simpler design. We simplified the task, which fared well in cognitive pretests. The number of attributes and levels is now well beyond the number where satisficing is a problem (Bansak et al., 2021). Second, we replicated our conjoint analyses using the "rating" variable instead of the "choice" variable (see Appendix G). Third, we conducted several standard robustness tests in Appendix H (carryover effects, profile order effects, screen size, speeding, choice task round). The tests indicate that the results shown above are robust.

Conclusion

Our analysis provides a partial answer to why welfare state recalibration is politically difficult despite the apparent popularity of social investment. Governments have little room for maneuver in times of austerity and find it impossible to cover all social risks equally. Consequently, they face substantial trade-offs when they have to design welfare state reforms that increase spending on some social policies at the expense of others.

While support for all social policies declines sharply when we account for such trade-offs, citizens have clear social policy priorities regarding welfare state recalibration. The split experiment shows that increases in education spending are more popular than increases in childcare spending and ALMP. The latter fail to reach majority support when trade-offs are involved. In particular, decreases in pension spending are politically difficult. According to our results, no trade-off that involves pension spending finds majority support.

The conjoint experiment provides for the first time evidence on citizen's multidimensional social policy priorities. We show that citizens attach a high priority to pensions and education. In contrast, childcare and child benefits only have a minor influence on the probability that respondents support a given reform, implying that they are of medium priority to citizens. Finally, active and passive labor market policies are unpopular in the context of austerity. When forced to choose, the average citizen does not support additional spending on the unemployed

if this comes at the cost of lower spending in other areas. Citizens thus attach a low priority to labor market policies.

Overall, we demonstrate that multidimensional welfare state recalibration towards social investment is an uphill battle because citizens do not strongly prioritize social investment over social consumption. Moreover, strong policy constituencies are likely to block many welfare state recalibration attempts, given that citizens react sensitively to changes in policies from which they benefit directly: Pensioners care more about increasing pension spending than working-age people, parents with young children prioritize family policies more strongly, and outsiders attach a greater priority to labor market policies. Therefore, the narrow self-interest of policy constituencies is a strong predictor of their social policy priorities and likely of their political behavior.

There are several important implications of our findings. From a methodological perspective, regular opinion polls consistently overstate the support for policies if respondents do not face the inherent policy trade-offs. We argue that such questions do not allow inferences about respondents' priorities towards welfare state recalibration. Our paper proposed two experimental strategies to present respondents with realistic trade-offs, which can easily be applied to other policy fields. Furthermore, our results show that the support structure for social investment and social consumption policies *varies* between and *within* these respective domains. Thus, it is not enough to use education and pensions as proxies for social investment and consumption, making it crucial to add more nuance to the literature on public opinion towards welfare state recalibration.

Substantively, our findings show that governments aiming for a social investment turn need to design welfare state reforms carefully. Political conflicts in mature welfare states are more about distributing resources to specific groups than about general support for the welfare state. Especially in times of austerity, welfare state recalibration will create winners and losers. The latter need to be compensated for welfare state reforms to gain political support. A key question for future research is how policy constituencies can be convinced to give up some privileges and support policies that address new social risks. In all likelihood, policymakers confronted with such electoral constraints will need to combine material incentives with solidaristic appeals. However, there is also some evidence that governments may be able to generate support for

welfare state retrenchment by controlling the political messaging: By framing cuts as fiscally necessary, political actors can sometimes avoid electoral punishment (Barnes and Hicks, 2018; Bisgaard and Slothuus, 2018). Future research should thus study how policymakers can use elite cues to make welfare state recalibration politically feasible despite the salient trade-offs emphasized here.

References

Ansell, B. W. (2010). *From the Ballot to the Blackboard: The Redistributive Political Economy of Education*. Cambridge: Cambridge University Press.

Armingeon, K. and G. Bonoli (2007). *The Politics of Post-Industrial Welfare States: Adapting Post-War Social Policies to New Social Risks*. London: Routledge.

Armingeon, K. and R. Bürgisser (2021). Trade-offs between redistribution and environmental protection: the role of information, ideology, and self-interest. *Journal of European Public Policy 28*(4), 489–509.

Bansak, K., J. Hainmueller, D. J. Hopkins, and T. Yamamoto (2020). Using conjoint experiments to analyze elections: The essential role of the average marginal component effect (AMCE). Available at SSRN: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3588941.

Bansak, K., J. Hainmueller, D. J. Hopkins, and T. Yamamoto (2021). Beyond the breaking point? survey satisficing in conjoint experiments. *Political Science Research and Methods 9*(1), 53–71.

Barnes, L. and T. Hicks (2018). Making austerity popular: The media and mass attitudes toward fiscal policy. *American Journal of Political Science 62*(2), 340–354.

Beramendi, P., S. Häusermann, H. Kitschelt, and H. Kriesi (2015). *The Politics of Advanced Capitalism*. Cambridge: Cambridge University Press.

Bisgaard, M. and R. Slothuus (2018). Partisan elites as culprits? how party cues shape partisan perceptual gaps. *American Journal of Political Science 62*(2), 456–469.

Blyth, M. (2013). *Austerity: The History of a Dangerous Idea*. Oxford: Oxford University Press.

Boeri, T., A. Börsch-Supan, and G. Tabellini (2001). Would you like to shrink the welfare state? a survey of european citizens. *Economic policy 16*(32), 08–50.

Bonoli, G. (2007). Time matters: Postindustrialization, new social risks, and welfare state adaptation in advanced industrial democracies. *Comparative Political Studies 40*(5), 495–520.

Bonoli, G. (2013). *The Origins of Active Social Policy: Labour Market and Childcare Policies in a Comparative Perspective*. Oxford: Oxford University Press.

Bouget, D., H. Frazer, E. Marlier, S. Sabato, and B. Vanhercke (2015). Social investment in europe: A study of national policies, european social policy network (espn). *Brussels: European Commission*.

Bremer, B. and R. Bürgisser (2021). Do citizens care about government debt? evidence from survey experiments on budgetary priorities. *European Journal of Political Research*.

Bürgisser, R. (2019). *The Politics of Welfare State Recalibration in Continental and Southern Europe*. Florence: European University Institute: PhD Thesis.

Busemeyer, M. R. and J. L. Garritzmann (2017). Public opinion on policy and budgetary trade-offs in european welfare states. *Journal of European Public Policy 24*(6), 871–889.

Busemeyer, M. R. and E. Neimanns (2017). Conflictive preferences towards social investments and transfers in mature welfare states. *Journal of European Social Policy 27*(3), 229–246.

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

Fossati, F. (2018). Who wants demanding active labour market policies? *Journal of Social Policy 47*(1), 77–97.

Fossati, F. and S. Häusermann (2014). Social policy preferences and party choice in the 2011 swiss elections. *Swiss Political Science Review 20*(4), 590–611.

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

Garritzmann, J. L., M. R. Busemeyer, and E. Neimanns (2018). Public demand for social investment: New supporting coalitions for welfare state reform in western europe? *Journal of European Public Policy 25*(6), 844–861.

Hainmueller, J. (2012). Entropy Balancing for Causal Effects. *Political Analysis 20*(1), 25–46.

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

Hall, P. A. and D. W. Soskice (2001). *Varieties of Capitalism: The Institutional Foundations of Comparative Advantage.* Oxford: Oxford University Press.

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

Häusermann, S. (2012). The Politics of Old and New Social Policies. In G. Bonoli and D. Natali (Eds.), *The Politics of the New Welfare State*, pp. 111–132. Oxford: Oxford University Press.

Häusermann, S., T. Kurer, and D. Traber (2018). The politics of trade-offs: Studying the dynamics of welfare state reform with conjoint experiments. *Comparative Political Studies 52*(7), 1059–1095.

Häusermann, S., M. Pinggera, M. Ares, and M. Enggist (2021). Class and social policy in the knowledge economy. *European Journal of Political Research.*

Hemerijck, A. (2013). *Changing Welfare States.* Oxford: Oxford University Press.

Hemerijck, A. (2017). *The Uses of Social Investment.* Oxford: Oxford University Press.

Hemerijck, A. (2018). Social investment as a policy paradigm. *Journal of European Public Policy 25*(6), 810–827.

Hoerl, A. E. and R. W. Kennard (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics 12*(1), 55–67.

Horiuchi, Y., D. M. Smith, and T. Yamamoto (2018). Measuring voters' multidimensional policy preferences with conjoint analysis. *Political Analysis*, 1–20.

Jensen, C. and M. B. Petersen (2017). The deservingness heuristic and the politics of health care. *American Journal of Political Science 61*(1), 68–83.

Leeper, T. J., S. B. Hobolt, and J. Tilley (2020). Measuring subgroup preferences in conjoint experiments. *Political Analysis 28*(2), 207–221.

Levy, J. D. (1999). Vice into virtue? Progressive politics and welfare reform in continental Europe. *Politics and Society 27*(2), 239–73.

Meltzer, A. H. and S. F. Richard (1981). A rational theory of the size of government. *Journal of Political Economy 89*(5), 914–927.

Morel, N., B. Palier, and J. Palme (2012). *Towards a Social Investment Welfare State? Ideas, Policies and Challenges.* Bristol: Policy Press.

Neimanns, E., M. R. Busemeyer, and J. L. Garritzmann (2018). How popular are social investment policies really? evidence from a survey experiment in eight western european countries.

European Sociological Review 34(3), 238–253.

Pierson, P. (1993). When effect becomes cause: Policy feedback and political change. *World Politics 45*(4), 595–628.

Pierson, P. (1996). The new politics of the welfare state. *World Politics 48*(2), 143–179.

Rehm, P. (2011). Social policy by popular demand. *World Politics 63*(02), 271–299.

Roosma, F., J. Gelissen, and W. v. Oorschot (2013). The multidimensionality of welfare state attitudes. *Social Indicators Research 113*(1), 235–255.

Rueda, D. (2007). *Social Democracy Inside Out. Partisanship and Labor Market Policy in Industrialized Democracies*. Oxford: Oxford University Press.

Schwander, H. (2020). Labor market insecurity among the middle class: a cross-pressured group. *Political Science Research and Methods 8*(2), 369–374.

Schwander, H. and S. Häusermann (2013). Who is in and who is out? A risk-based conceptualization of insiders and outsiders. *Journal of European Social Policy 23*(3), 248–269.

Seeleib-Kaiser, M. (2008). *Multiple and Multi-Dimensional Welfare State Transformations*. London: Palgrave Macmillan UK.

Taylor-Gooby, P. (2004). *New Risks, New Welfare*. Oxford: Oxford University Press.

van Oorschot, W. (2006). Making the difference in social europe: Deservingness perceptions among citizens of european welfare states. *Journal of European Social Policy 16*(1), 23–42.

Vlandas, T. (2020). The political consequences of labor market dualization: Labor market status, occupational unemployment and policy preferences. *Political Science Research and Methods 8*(2), 362–368.

Zucman, G. (2015). *The Hidden Wealth of Nations: The Scourge of Tax Havens*. University of Chicago Press.

Supplementary Information/Online Appendix

Public Opinion on Welfare State Recalibration in Times of Austerity: Evidence from Survey Experiments

December 2021

Table of Contents

A Operationalization and summary statistics of key variables

Table A.1: Operationalization of variables used as covariates

Variable	Question Wording	Measurement
Age	What year were you born?	Numerical variable: 18-88
Female	What is your sex?	Binary: 1 = female; 0 = men or no answer
Married	What is your marital status?	Binary: 1 = married; 0 = not married
Children	How many children are living in your household under the age of 10?	Binary: 1 = children; = no children
Education	What is the highest level of education that you have achieved so far?	Categorical: 1 = primary / lower secondary; 2 = secondary; 3 = tertiary
Class*	What is/was your occupation?	Categorical: 1 = employer; 2 = middle class; 3 = working class; 4 = routine worker
Income	Please indicate the answer that includes your total household income in the previous month, after taxes and compulsory deductions?	Numerical variable by percentile: 1-10 OR categorical variable by percentile: 1 = 1st - 3rd; 2 = 4th - 7th; 3 = 8th - 10th
Retired	Which of these descriptions applies to what you have been doing for the last months?	Binary variable: 1 = retired; 0 = otherwise
Work contract	Based on three questions: (i) What is/was your occupation? (ii) What is/was the nature of your employment? (iii) What describes best your current working status?	Categorical: 1 = outsiders (part-time workers with < 30 hours, temporary workers, or unemployed); 2 = insiders (full-time permanent working contract); 3 = upscales (higher skilled professionals, large employers & business owners, and self-employed citizens); 4 = out of work (retired, homemakers, students)
Union	Are you or have you ever been a member of a trade union or similar organization?	Binary: 1 = current union member; 0 = otherwise
Partisanship*	Which party did you vote for in the last (COUNTRY) election in (MONTH/YEAR)?	Categorical: 1 = far left; 2 = center left; 3 = center right; 4 = far right; 5 = other party; 0 = abstention
Left-right scale	In politics people sometimes talk of "left" and "right". Using the scale below, where would you place yourself, where 0 means the left and 10 means the right?	Categorical: 0-4 = left; 5-6 = center; 7-10 = right

* = Please see below for further information on the exact operationalization.

Class was coded according to the following occupational groups. (1) *Small and large employers*: self-employed farmers, fishermen, professionals, and shop owners/ business proprietor. (2) *Middle class*: employed professionals, general management, middle management, employed position desk with upper secondary education, employed position travelling with upper secondary education, and employed position service with upper secondary. (3) *Working class*: skilled manual workers, supervisors, and unskilled workers with an upper secondary education. (4) *Routine workers*: unskilled worker without upper secondary, employed position desk without upper secondary education, employed position travelling without upper secondary education, and employed position service without upper secondary

Partisanship was coded according to the ParlGov database, which classifies parties into families by their position in an economic (state/market) and a cultural (liberty/authority) left/right dimension. The party families were then simplified into five political groups, as shown below:

Table A.2: Classification of political parties into five groups

	Germany	**Italy**	**United Kingdom**
Far right	Alternative für Deutschland	Lega Nord, Fratelli d'Italia	UK Independence Party
Center right	Christlich Demokratische Union, Christlich-Soziale Union, Freie Demokratische Partei	Scelta Civica, Nuovo Centrodestra, Forza Italia	Conservative, Liberal Democrats
Center left	Sozialdemokratische Partei Deutschlands, Grüne/Bündnis90	Partito Democratico, Radicali Italiani, Articolo 1/MDP	Labour, Scottish National Party, Greens
Far left	Die Linke	Sinistra Italiana, Movimento 5 Stelle, Rifondazione Comunista	
Others	Piratenpartei	Südtiroler Volkspartei	Plaid Cymru

Table A.3: Summary Statistics

Variable	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
Education spending (pooled)	3,600	6.309	2.602	0	5	8	10
Childcare spending (pooled	3,600	5.530	2.725	0	4	7	10
ALMP spending (pooled)	3,600	5.418	2.715	0	4	7	10
Split	3,600	2.498	1.117	1	1.8	3	4
Age	3,600	46.883	16.029	18	34	60	87
Female	3,597	0.513	0.500	0.000	0.000	1.000	1.000
Married	3,600	0.459	0.498	0	0	1	1
children_u10	3,600	0.209	0.407	0	0	0	1
Education_c	3,585	2.213	0.676	1.000	2.000	3.000	3.000
Class	2,999	2.186	0.826	1.000	2.000	3.000	4.000
Income	3,427	4.967	2.803	1.000	2.000	7.000	10.000
Retired	3,600	0.207	0.405	0	0	0	1
Union	3,546	0.272	0.445	0.000	0.000	1.000	1.000
Partisanship	3,335	2.163	1.302	0.000	1.000	3.000	5.000
Left-right scale	3,600	1.845	0.737	1	1	2	3
Work scale	3,289	2.534	1.067	1.000	2.000	3.000	4.000

B Case selection, sampling, and quality tests

B.1 Case selection

Table A.4: Case Selection

	DE	IT	UK
Welfare regime	Continental	Southern	Liberal
Varieties of capitalism	CME	MME	LME
Unemployment rate, 2017	3.7	11.2	4.3
Youth unemployment rate, 2017	6.8	34.8	12.1
Government debt (% of GDP), 2017	72	153	117
Education spending (% of GDP), 2016	3.6	3.3	4.2
Pension spending (% of GDP), 2016	10.1	16.2	6.2

B.2 Sampling

Our survey was fielded in three large European countries in January 2018: Germany, Italy, and the United Kingdom. In each country, 1,200 respondents were recruited to participate in the survey. A large online panel provided by *Qualtrics* was used with tens of thousands of panelists across all age brackets. Qualtrics only uses double-opt in online panels that allow their panelists to take a survey no more than once every two weeks. Respondents were then drawn from a pool of eligible voters in each country and the sample was representative of all eligible voters based on gender and age, meaning we introduced quota sampling based on age and gender.

Given our subgroups comparisons, we are particularly interested in having roughly representative income and partisan groups. Even without quotas sampling by income and ideology, our sample is already fairly representative in terms of both variables. Figure A.1 plots the share of respondents by income decile for all countries (left) and by country (right). Overall, our income variable is almost equally distributed across income decile, only with a slightly higher representation of low-income respondents compared to high-income respondents. While the German sample is quite well balanced across income decile, the Italian and British sample is slightly more skewed towards low-income respondents. However, the differences are overall not particularly large.

In addition, we compared the party vote share in the closest election to the share of voters in our survey (Table A.5). We distinguish between the far right, center right, center left, and the far left, as explained in more detail in Table A.2. Especially in Italy, our sample matches the actual vote share that these four blocks received very well. In Germany, far left and center left voters are slightly over-represented, while center right voters are underrepresented. In the United Kingdom, center right voters are also underrepresented while center left and far right voters are slightly over-represented. Overall, the survey vote share matches fairly well the actual vote share.
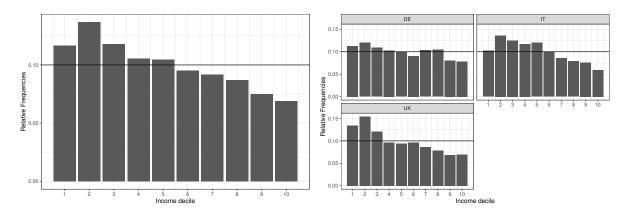
Figure A.1: Share of respondents by income decile (left: all countries; right: by country)

Table A.5: Survey vote share versus actual vote share

		Far Left	Center Left	Center Right	Far Right
DE	Election	9.2	29.4	43.7	12.6
	Survey	15.9	32.1	33.2	11.2
	Diff.	6.7	2.7	-10.5	-1.4
IT	Election	34.4	22.2	18.8	21.8
	Survey	38.5	22.1	15.6	19.9
	Diff.	4.1	-0.1	-3.2	-1.9
UK	Election	0.0	47.0	49.9	1.8
	Survey	0.0	51.6	40.9	5.7
	Diff.	0.0	4.6	-9.0	3.9

We also used weights to match the demographic characteristics of the country's population as closely as possible using entropy balancing (Hainmueller 2012). Please see Appendix F for a detailed explanation and the results using entropy balancing.

B.3 Quality checks and screening out of respondents

To ensure the quality of our sample, the survey included an attention check and a speeding check. We implemented a common attention check in the middle of the survey. Respondents were shown the following question:

We are interested in learning about your preferences on a variety of topics, including colours. To demonstrate that you have read this question, please go ahead and select both red and green from the alternatives below, no matter what your favourite color is. Yes, ignore the question below and select both of those options. What is your favourite color?

The answer options were: blue, red, purple, orange, yellow, green, brown, gray. Those who did not select the two correct color, were excluded from the survey. Additionally, we implemented a speeding check for the whole survey and one specifically for the conjoint experiment. Respondents who had a response time below 1/3 of the median response time were excluded.

C   Instructions for the Conjoint Experiment

The full instructions for the conjoint tasks are shown below. First, respondents were presented with the following introduction to the experiment:

Please take your time and read the information below very carefully. It contains the instructions for the next part of the survey.

Every year the [COUNTRY] government spends money in a variety of different areas. We are interested in how you would like the government to change its spending pattern.

We will now show you several proposals for possible changes to government spending in different areas. We will always show you two possible proposals in comparison. For each comparison we would like to know which of the two proposals you prefer. You may like both proposals or neither. In any, case please choose the proposal that you like the most. In total, we will show you five comparisons.

The possible proposals only include changes with regard to a few selected types of government spending. Please assume that spending in all other areas is held constant. Please also assume that taxation and the level of government debt are held constant.

People have different opinions about this issue and there are no right or wrong answers. Please always take your time when reading the proposals.

After this, the screen presented two reform proposals, as shown in Figure A.2. Respondents were asked five times to choose (i) between two packages ("choice variable") and (ii) to indicate how likely they are to support each of the proposals ("ranking variable").

Please carefully review the options detailed below, then please answer the questions.

Which of these proposals do you prefer?

	Proposal 1	Proposal 2
Child benefits	Decrease spending	No change
Old-age pensions	No change	Decrease spending
Education	No change	Increase spending
Training for the unemployed	Increase spending	No change
Unemployment benefits	Increase spending	Decrease spending
Childcare services	Decrease spending	Increase spending

Proposal 1

Proposal 2

How would you rate proposal 1 on a scale from 0 to 10, where 0 indicates that the government should definitely not adopt the proposal and 10 indicates that the government should definitely adopt it?

0 - Definitely not adopt	1	2	3	4	5	6	7	8	9	10 - Definitely adopt

How would you rate proposal 2?

0 - Definitely not adopt	1	2	3	4	5	6	7	8	9	10 - Definitely adopt

Figure A.2: Screenshot of a conjoint task presented to respondents

D Explanation of ridge regression to analyze the conjoint experiment

We propose ridge regression as a novel method to analyse conjoint survey experiments with dependencies. In our design, the values that each attribute can take are linearly dependent on the other attributes to ensure that the budget is balanced. Therefore, there is explicit confounding between the features of the fiscal packages, due to the restrictions to the randomization protocol that we introduced. We included these restrictions to increase external validity, as recommended by Hainmueller et al. (2014, p. 26). They argued that "researchers should employ restrictions and exclude attribute combinations whenever they are deemed so unrealistic that a counterfactual would essentially be meaningless." We took this recommendation to the extreme because the budget constraint of the public budget is binding: changing the government's expenditure or revenue on one dimension necessarily has consequences for another dimension. This poses difficulties for traditional regression analysis, but it ensures external validity.

To see this, imagine the alternative: We could have introduced combinations that are unrealistic to break the super-collinearity that exists in reality. This would have allowed us isolate the effect of each attribute independently of other attributes, but it poses a threat to external validity. Respondents would have probably most preferred atypical profiles, which increase government spending, decrease taxation, and cut government debt all at the same time. Given our interest in citizen's priorities in the face of hard budgetary tade-offs, this would not have been satisfactory. As Bansak et al. (2020, p. 24-25) argue, "the AMCE averages the effect of an attribute over two different distributions: the randomization distribution of the other attributes and the distribution of respondents." The AMCEs of a fully randomized design, therefore, are not of interest to us. We only wanted to make inferences about realistic budgetary combinations, i.e., those that are fully balanced.

The resulting experimental design necessitates a modelling approach which accounts for the design-based super collinearity. One common solution to enable OLS standard regression analysis in instances of super collinearity is to drop one of the correlated variable. This strategy usually works well, but it is not useful in our case because it defeats the point of the design. We are interested in the support for a fiscal package as a function of all of its individual attributes. Moreover, dropping one attribute from the analysis may lead to specification bias.

We, therefore, propose to use regularization to analyse the results from conjoint experiments, which are perfectly multicollinear. Specifically, we use ridge regression, which is a common regularization method that adds a penalty term to the common OLS regression. Hoerl (1962) and Hoerl and Kennard (1970) suggested to use a ridge regression as an ad-hoc fix to address instances of high multicollineary, including instances of design-based collinearity. It allows one to estimate coefficients for all independent variables in the model even in the presence of super collinearity, and consequently, the method is also used in fields such as genetics where a set of related genetic predictors may jointly cause certain diseases.

To see how ridge regression works, recall that OLS regression attempts to minimize the sum of errors squared, as shown in the equation below:

$$\sum_{i=1}^{M}(y_i - \hat{y}_i)^2 = \sum_{i=0}^{M}(y_i - \sum_{j=0}^{p} w_j w_{ij})^2 \tag{1}$$

Ridge regression adds the following term to this model:

$$+ \lambda \sum_{j=0}^{p} w_j^0 \tag{2}$$

This term is referred to as the ridge penalty and λ is the penalty parameter. If λ is zero, then ridge regression is essentially an OLS regression. However, if λ is above zero, then it adds a constraint to the coefficient. This constraint minimizes the coefficients (which is called shrinkage), which results in a lower variance and a lower error value. Consequently, ridge regression is a way to decrease the complexity of a model without reducing the number of variables. It is a solution to a constrained estimation problem.

Importantly, the ridge penalty shrinks large regression coefficients of correlated predictors and reduces overfitting. Contrary to Lasso regression, an alternative regularization method, ridge regression does not shrink coefficients to zero.[8] It includes all independent variables in the data and is thus a good way to analyze results from conjoint survey experiments with a large number of restrictions. The "shrinkage estimators" performs better than OLS when the data matrix is relatively sparse, but it introduces a bias in the estimates due to the ridge penalty. Given that the bias introduced by the ridge penalty is systematic, it still allows us to make inferences about respondents' priorities in our case. Our empirical strategy thus provides a modelling strategy for the underlying utility function behind respondents' choice of fiscal packages, while maintaining predictive performance and interpretability.

We use the R package `glmnet` to find the best value of λ through cross-validation. We then proceed by using this optimal λ to estimate the regression coefficients (also referred to as AMCEs) and marginal means. Ridge regression does not provide standard errors for coefficients, but we rely on non-parametric boot-strapping to calculate standard errors and confidence intervals. To this end, we wrote a R function which calculates the same ridge regression 1000 times with a random sample of our observations and calculates standard errors based on the uncertainty of the results. This is a popular method for parametric inference, and it allows us to assign a measure of accuracy to the coefficients obtained from the ridge regression.

---

[8]Lasso regression is thus especially useful for model selection.
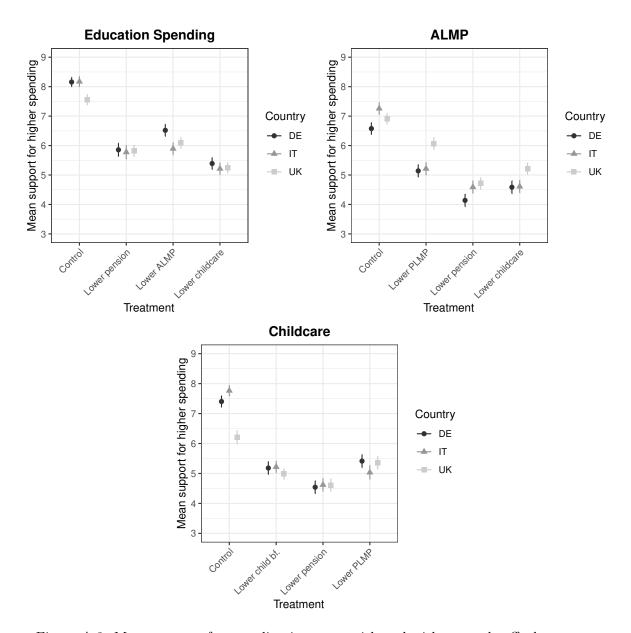
E   Analysis by country

Figure A.3: Mean support for spending increases with and without trade-offs, by country
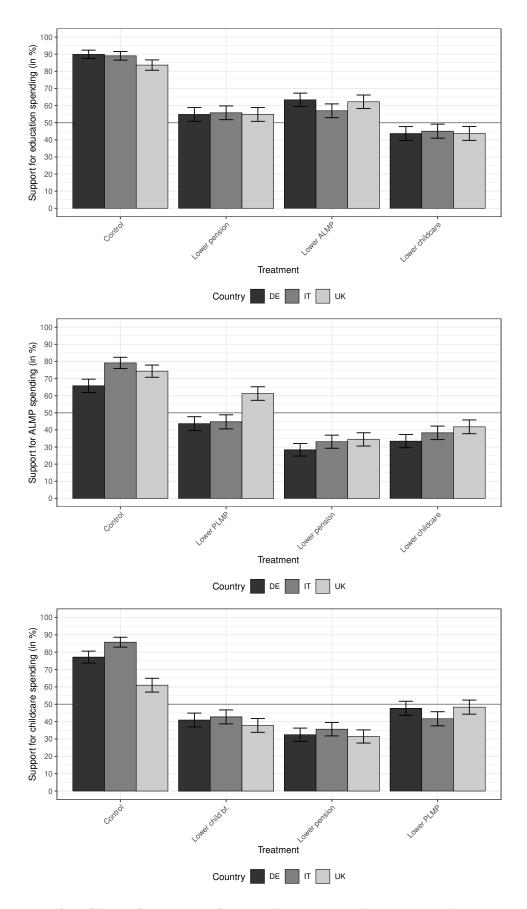
Figure A.4: Share of supporters for spending increases by treatment, by country
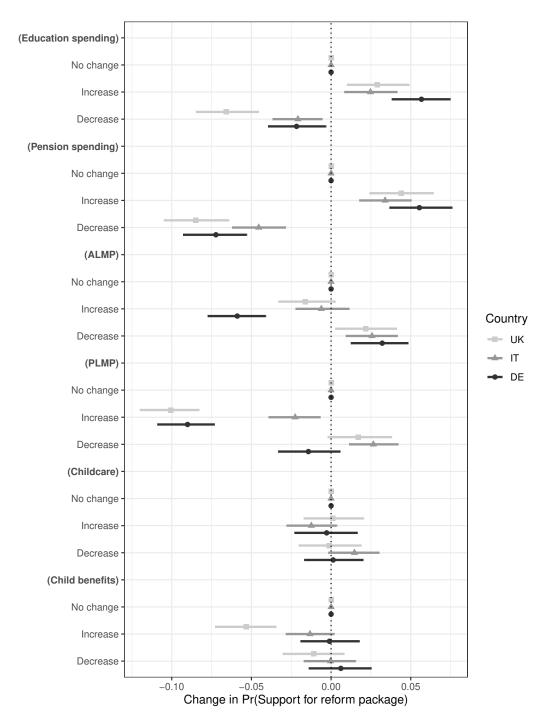
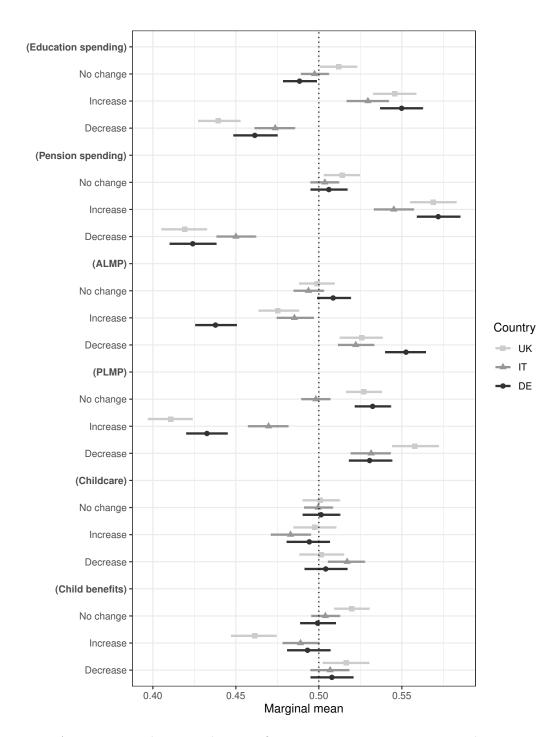Figure A.5: AMCEs from conjoint survey experiment by country

Figure A.6: Estimated marginal means from conjoint survey experiment by country

F Analysis with weights created by entropy balancing

We worked with the survey company Qualtrics to obtain an online sample that was representative of the population based on key indicators such as age and gender in each country. To test that there were no other biases in our sample that influence our results, however, we used entropy balancing to create survey weights using the R package `ebal`. Based on margins from the population in each country, we created these weights based on age, gender, education, and economic activity and replicated our analyses. The results are not substantively different from the results shown in the main text.
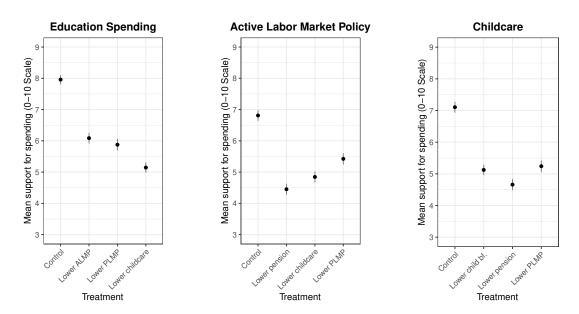
Figure A.7: Average support for spending increases by treatment with survey weights, pooled
Note: The figure shows the mean support and 95 percent confidence intervals for different policies without trade-offs (control group) as well as with the three respective trade-offs (treatment groups). Mean support is measured on a scale from 0 to 10.
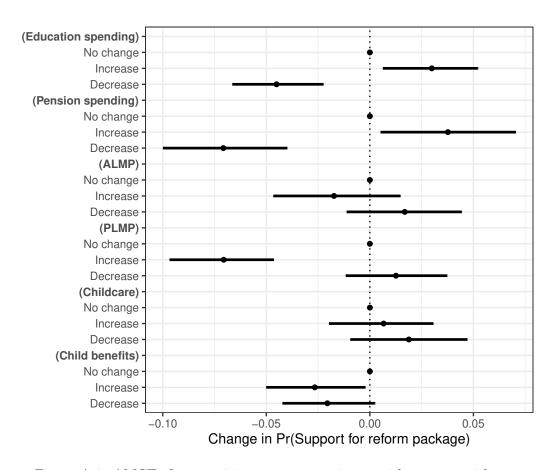
Figure A.8: AMCEs from conjoint survey experiment with survey weights

G Additional results from the conjoint survey experiment

We replicate our analysis with a second dependent variable, namely the rating variable. For this variable, respondents evaluated each reform proposal that they saw on a scale from 0 to 10. The results are very similar to the findings presented in the main text. We then show the full conjoint graphs for the difference between constituencies and non-constituencies shown in Figure 5. Moreover, to test the robustness of our findings, we further run several tests for heterogeneous effects in our conjoint survey experiment. Specifically, we tested for heterogeneous effects along for subgroups by ideology (left-right self placement), education level, and income. The results are shown below and they indicate that the heterogeneous effects are relatively small, except for specific groups on specific attributes (see our discussion on policy constituencies in the main text). This is also true for other possible subgroups (e.g., employment activity, gender, age) which are not shown below.

G.1 Results from the analysis with the rating variable
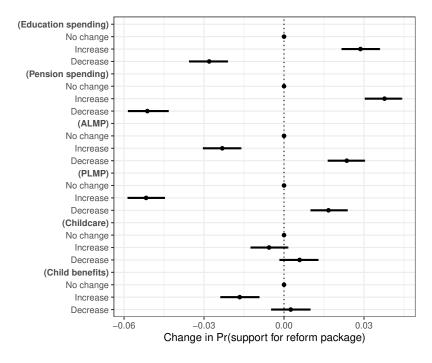
Figure A.9: AMCEs from conjoint survey experiment with the rating variable, pooled
Note: The figure shows the average component-specific marginal effect (ACME) of a change in the value of one of our six dimensions on the support for the reform package.

G.2   Full constituency plots showing heterogeneous effects
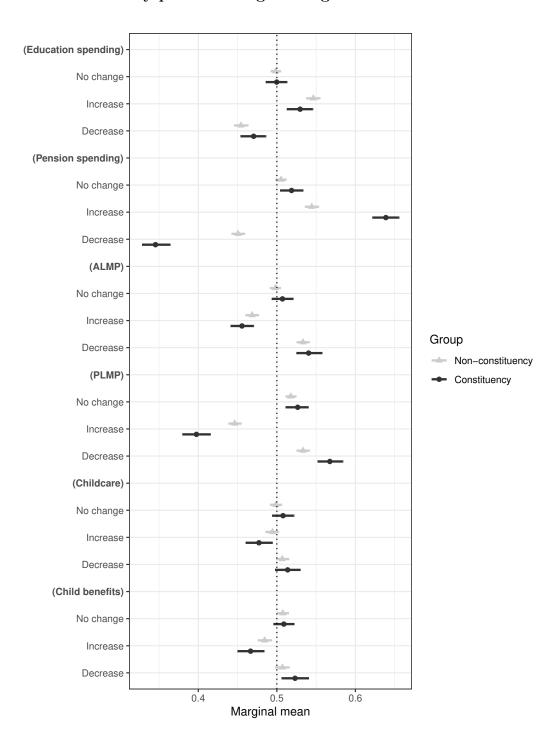
Figure A.10: Estimated marginal means from conjoint survey experiment for retired respondents vs. and all other respondents
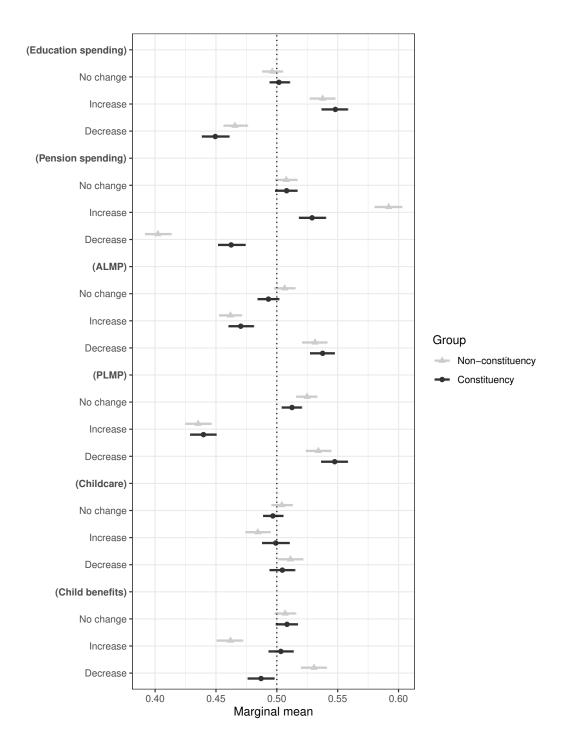
Figure A.11: Estimated marginal means from conjoint survey experiment for education constituency vs. all other respondents
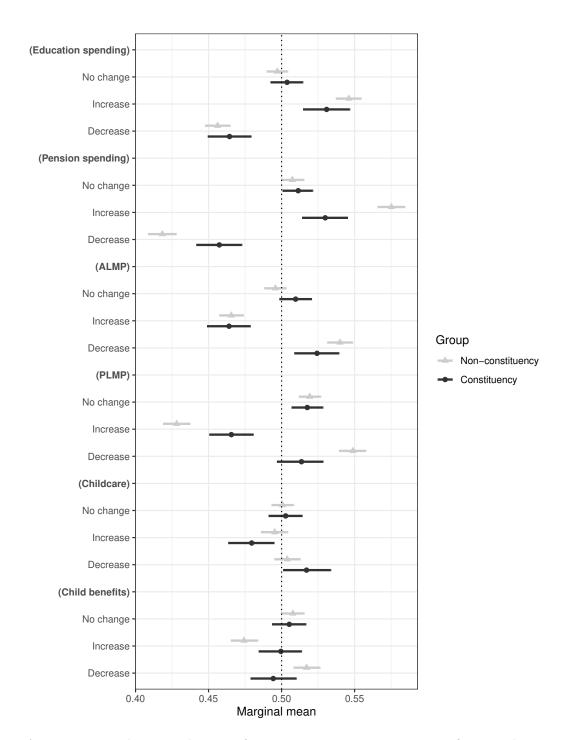
Figure A.12: Estimated marginal means from conjoint survey experiment for outsiders vs. all other respondents
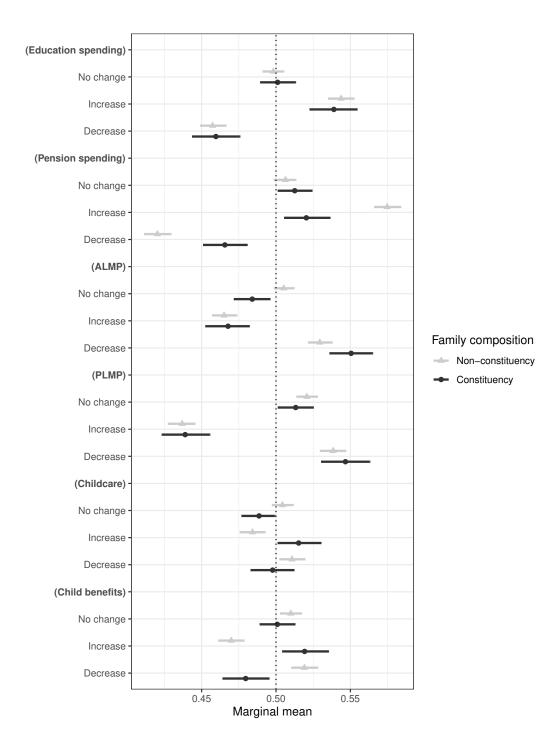
Figure A.13: Estimated marginal means from conjoint survey experiment for parents of children (under the age of 10) vs. all other respondents

G.3   Additional tests for heterogeneous treatment effects
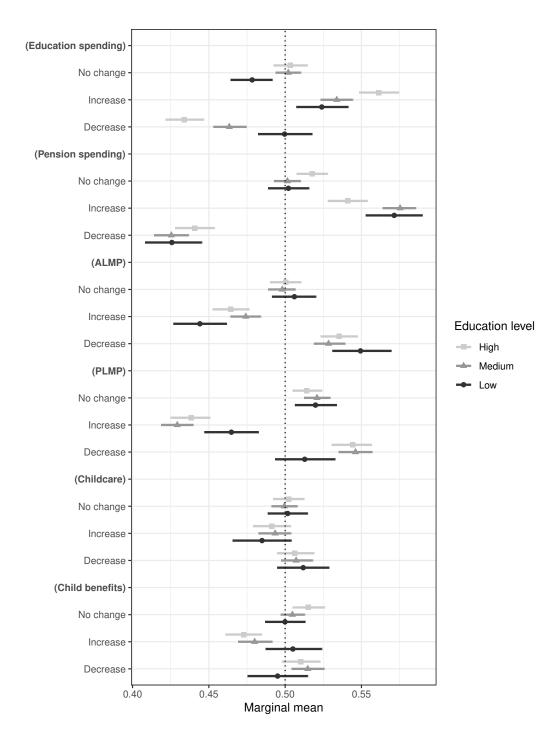
Figure A.14: Estimated marginal means from conjoint survey experiment by education
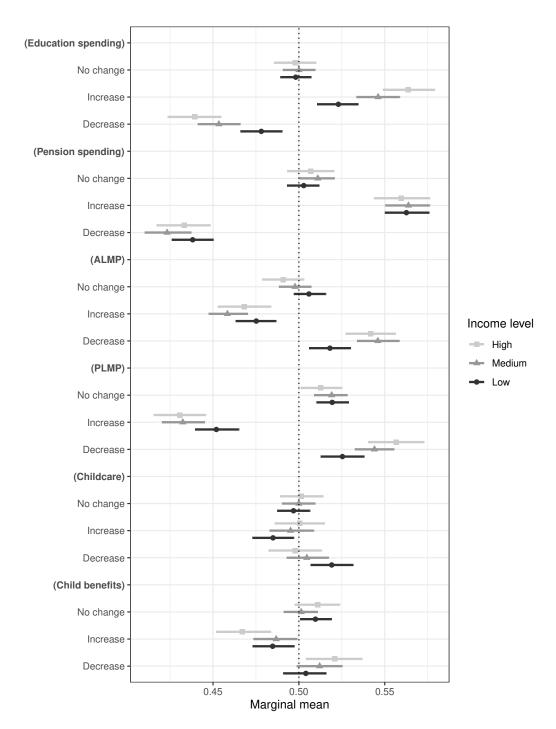
Figure A.15: Estimated marginal means from conjoint survey experiment by income
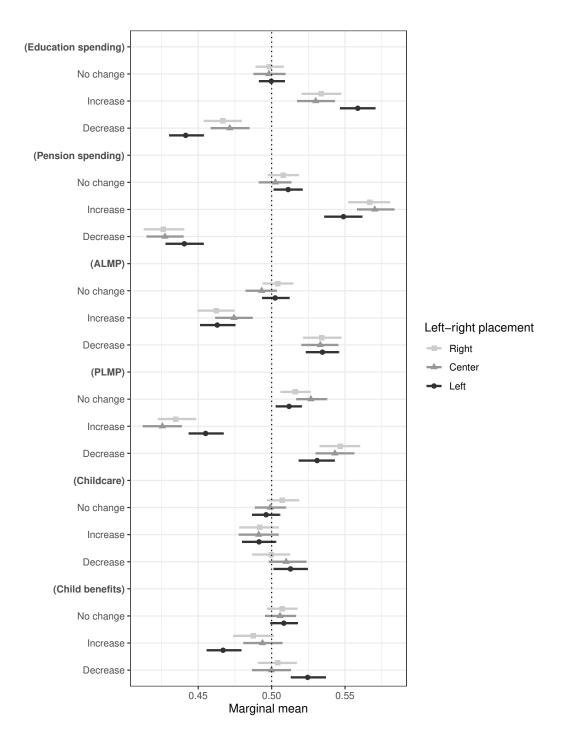
Figure A.16: Estimated marginal means from conjoint survey experiment by left-right placement

H   Robustness tests for the conjoint survey experiment

We have used a series of tests to check whether the results are robust. They were designed to check that the common assumptions involved in conjoint analysis are satisfied and to probe potential concerns about the validity of the results.

On the one hand, we conducted the diagnostic tests suggested by Hainmueller et al. (2014). First, conjoint analyses relies on the assumption that there are no carryover effects between the different rounds of conjoint tasks. To test whether this assumption holds, we estimated the AMCEs separately for each of the five rounds of conjoint tasks. The results are shown in figures A.17.[9] Second, we checked whether there are profile order effects, i.e., whether the AMCEs and marginal means depend on whether the attribute occurs in the first or second profile in a given task. To this end, we estimated AMCEs separately for all the observations where attribute levels occurred in the first and the second profile respectively (Figure A.18). Finally, note that we already addressed the concern about atypical profiles raised by Hainmueller et al. (2014) in the research design. Specifically, we included a large number of restrictions to prevent profiles that are unrealistic and would not occur in the real world.

On the other hand, we also used further robustness tests, which are important due to the design of the survey. First, we checked whether respondents lost concentration throughout the survey by estimating all results based on the first two (out of five) conjoint comparisons only. Moreover, we included round or task fixed effects to take account of the fact that respondents might make different choices in later stages of the conjoint experiment, for example due to fatigue or lack of concentration. The results for the first two (out of five) conjoint comparisons are shown in Figure A.19.

Second, we assessed the relative time that respondents took to complete the conjoint tasks and we excluded those respondents that speed through the conjoint tasks, comparing the results with the overall sample. We also distinguished respondents by the time that they took overall for the survey and used subgroup analysis to test whether our results are robust across groups. The results are shown in Figure A.20.

Third, the conjoint survey experiment described above was embedded in a survey, which included two different set of conjoint tasks. The order in which these conjoint experiments occurs in the survey was randomized. Still, we checked whether respondents are influenced in their evaluations of the conjoint profiles if they have already completed a different set of conjoint tasks beforehand. For this purpose, we split the sample and analyzed the results separately depending on whether the conjoint experiment occurred before or after the other conjoint experiment in the survey (Figures A.21).

---

[9]Following, the suggestion by Leeper et al. (2020), we also estimated the conditional marginal means for each of the five rounds of conjoint tasks but the results are not shown.

Fourth, there is also a possibility that the screen size might affect the way respondents evaluate the conjoint tasks. We therefore also separately analyzed responses from mobile versus non-mobile respondents and checked to what extent they differ.
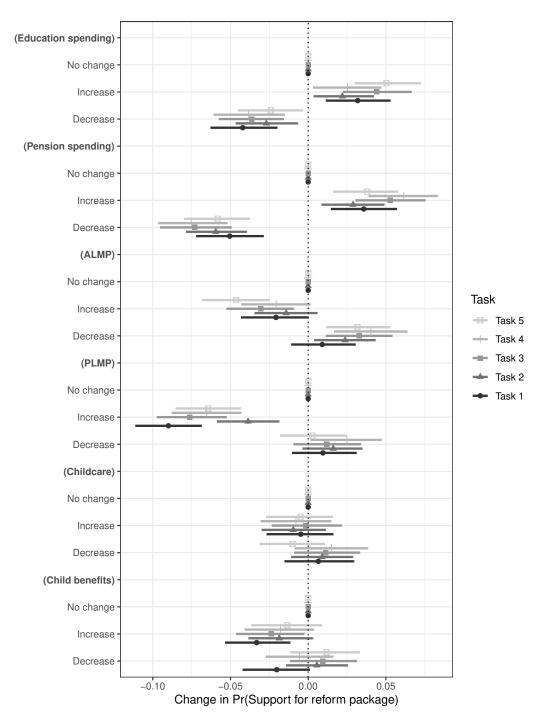

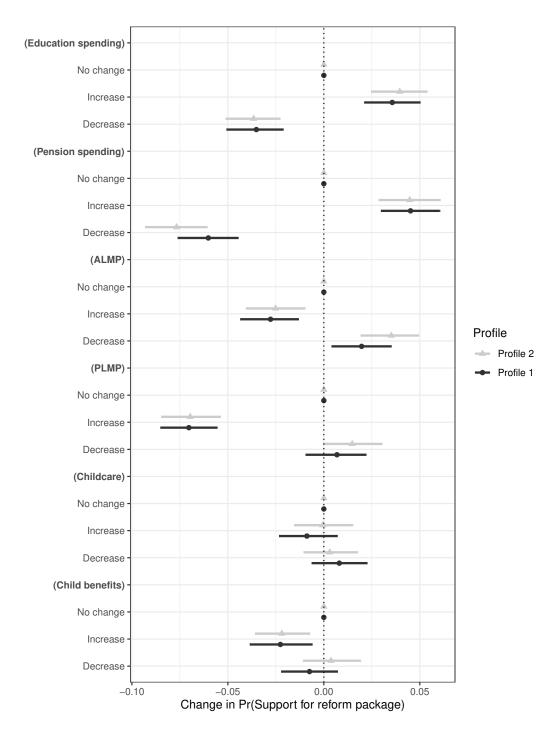Figure A.17: AMCEs from conjoint survey experiment by task

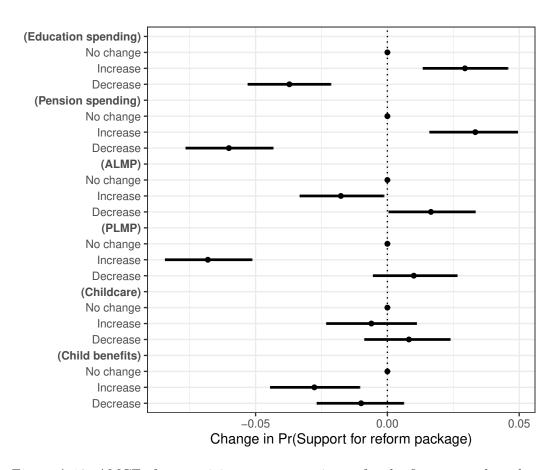Figure A.18: AMCEs from conjoint survey experiment by profile order

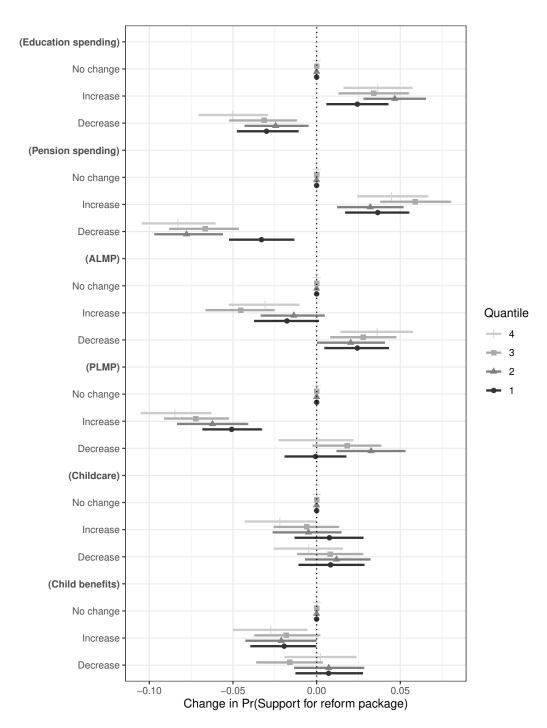Figure A.19: AMCEs from conjoint survey experiment for the first two tasks only

Figure A.20: AMCEs from conjoint survey experiment by speed of respondents
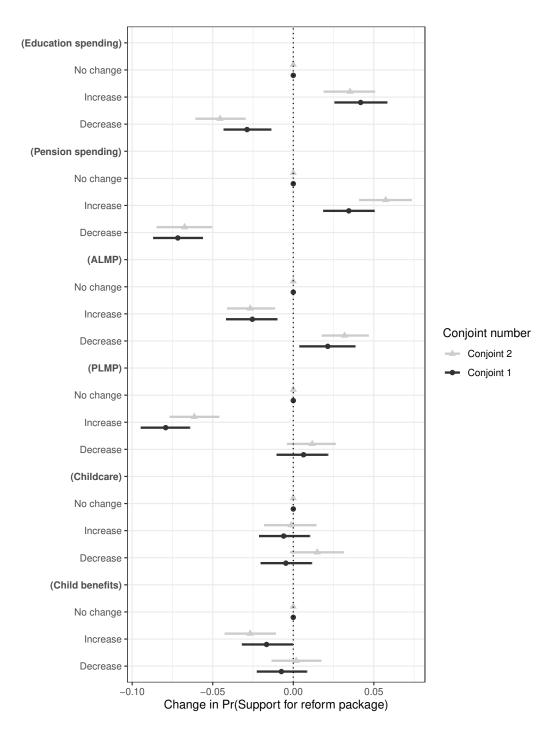
Figure A.21: AMCEs from conjoint survey experiment by order of the conjoint task in the survey
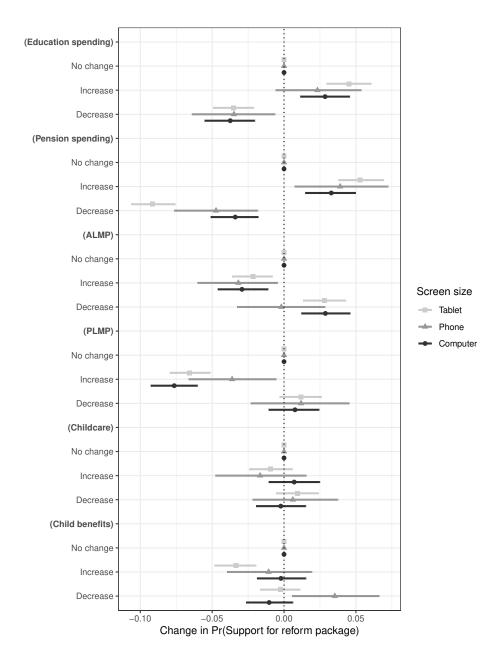
Figure A.22: Estimated marginal means from conjoint survey experiment by screen size of respondent's device