CESifo WORKING PAPERS

7276
2018

September 2018

Complex Ballot Propositions, Individual Voting Behavior, and Status quo Bias

Zohal Hessami, Sven Resnjanskij

CESifo

Impressum:

CESifo Working Papers
ISSN 2364-1428 (electronic version)
Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH
The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute
Poschingerstr. 5, 81679 Munich, Germany
Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email office@cesifo.de
Editors: Clemens Fuest, Oliver Falck, Jasmin Gröschl
www.cesifo-group.org/wp

An electronic version of the paper may be downloaded
· from the SSRN website: www.SSRN.com
· from the RePEc website: www.RePEc.org
· from the CESifo website: www.CESifo-group.org/wp

CESifo *Working Paper No. 7276*
Category 2: Public Choice

Complex Ballot Propositions, Individual Voting Behavior, and Status quo Bias

Abstract

One concern about direct democracy is that citizens may not be sufficiently competent to decide about complex policies. This may lead to exaggerated conservatism in the voting decision (status quo bias). To investigate how complexity affects individual voting behavior, we develop a novel measure of proposition complexity (using official pre-referendum booklets) and combine it with post-referendum survey data from Switzerland. Using Heckman selection estimations to account for endogenous variation in participation rates, we find that an increase in proposition complexity from the 10th to the 90th percentile would decrease voters' approval by 5.6 ppts, which is often decisive: an additional 12% of the propositions in our sample would be rejected.

JEL-Codes: D720, D780.

Keywords: voting behavior, proposition complexity, direct democracy, status quo bias, Heckman probit model.

*Zohal Hessami**
Department of Economics
University of Konstanz
Germany – 78457 Konstanz
zohal.hessami@uni-konstanz.de

Sven Resnjanskij
Ifo Institute – Leibniz Institute for Economic Research
at the University of Munich
Germany – 81679 Munich
resnjanskij@ifo.de

*corresponding author

The authors thank Kira Gährken, Sebastian Kopf, Marcus Rösch, Christoph Siemroth and Tobias Steinhauser for excellent research assistance and gratefully acknowledge funding by the German National Science Foundation (DFG) through Collaborative Research Center 884 ("Political Economy of Reforms") at the University of Mannheim. The authors also thank the Young Scholar Fund at the University of Konstanz for funding the follow-up project "The complexity of direct-democratic propositions, voter preferences, and the feasibility of political reforms". This article has benefitted from comments and suggestions by participants at the 2013 Workshop on Political Economy at CESifo in Dresden, the 2014 EEA meeting in Toulouse, the 2014 IIPF meeting in Lugano, the 2015 Silvaplana Workshop on Political Economy, the 2015 "Political economy: theory meets empirics" workshop in Konstanz, a political economy workshop at BI Norwegian Business School in 2016, a direct democracy workshop at DIW Berlin in 2016, the 2017 EPCS meeting in Budapest and various seminars in Konstanz.

I. Introduction

Direct legislation provides citizens with the opportunity to directly choose policies. This influence comes at a cost, however. Voters have to decide about propositions on a broad variety of issues. Often several propositions are at stake on the same day (i.e. on the same ballot) where citizens are asked to choose between rejecting a proposition (status quo) and accepting a proposition.[2] Since "a defining characteristic of many propositions is complexity" (Lupia 1994, 63), voters may face considerable difficulties in estimating the personal consequences of accepting or rejecting a proposition. Various authors even go so far to assert that ordinary voters may not be sufficiently competent to decide about complex policy issues (Cronin 1999; Magleby 1984).[3]

If voters do face such difficulties in deciding about complex propositions, how may this affect their voting behavior? First, it is likely that complexity has a negative influence on the individual decision to participate in a referendum. Voters may for instance derive less utility from turning out when they are uncertain about their voting decision (Matsusaka 1995). Second, the theoretical literature suggests that complexity increases the likelihood that voters reject a proposition. Samuelson and Zeckhauser (1988) were among the first to state that the existence of uncertainty – for instance due to the complexity and a resulting poor understanding of an issue – may lead to a status quo bias. Related to this, Eichenberger and Serna (1996) argue that complexity increases the likelihood that voters make random errors in assessing the costs and benefits of a proposed ballot measure. Due to the nature of the voting procedure these random errors have asymmetric effects on the outcome of a referendum and may lead to a higher likelihood of rejecting a proposition. Finally, a behavioral literature emphasizes that heuristics are primarily used when people have to trade off the effort required in decision-making and the accuracy of one's decision (Gigerenzer and Gaissmaier 2010). Hence, voters are particularly likely to rely on the status quo heuristic when they face highly complex propositions.

To be able to empirically investigate whether the complexity of propositions affects voters' decisions to turn out in a referendum and to accept or reject a proposition, we have to address two challenges. First, we need to find a way to measure proposition complexity. Most of the existing empirical literature on complexity in direct legislation resorts to a convenient

[2] Crowded ballots (i. e. ballots with many propositions) have been subject to criticism for many decades (see for instance Lapalombara and Hagan (1951)).

[3] In an influential study, Magleby (1984) examines California ballots during the 1970s and argues that more than 17 years of formal education would be required to understand an average proposition as stated on a ballot.

measure of *ballot complexity* which counts the number of propositions that were at stake on the same day (Selb 2008; Stadelmann and Torgler 2013).[4] Alternatively, some studies use the number of words or lines devoted to a proposition on a ballot as a measure of *proposition prolixity* (Bowler, Donovan, and Happ 1992; Kriesi 2008; Nicholson 2003).[5] The ballot complexity measure only takes into account how many propositions voters have to deal with on the same day (i.e. how crowded a ballot is) and thus does not capture the complexity of an individual proposition based on its content. On the other hand, the proposition prolixity measure refers to a point in time where a voter has already taken a first decision: she has turned out and is sitting in front of the ballot. Thus, using the prolixity measure one would ignore citizens who abstain in the first place. Second, one has to think carefully about how to capture the effect of complexity on individual voting behavior. While at first sight it may seem that the participation decision (whether to turn out) is irrelevant to the effect of proposition complexity on the referendum outcome and a potential status quo bias, this is likely not the case. Theoretical contributions such as the swing voter's curse theory by Feddersen and Pesendorfer (1996) underline the importance of potential interrelations between the participation and voting decision. Therefore, one needs (i) a dataset that includes information on individual participation and voting decisions and (ii) a suitable econometric specification that takes into account that the two decisions are interdependent.

In this paper, we address both of these challenges. Our dataset covers 223 federal referendums that were held in Switzerland between 1981 and 2010. The data on individual participation and voting decisions is taken from the VOX post-referendum survey series. We combine this rich micro data-set with a novel measure of proposition complexity which we construct based on information provided in official pre-referendum booklets that are sent to all Swiss households prior to a referendum. Our regression model takes into account that the participation and voting decisions are interrelated.[6]

---

[4] A higher number of propositions per ballot has been found to be positively associated with (i) a lower awareness of the propositions at stake (Nicholson 2003; Kriesi 2008), (ii) an interference with the pre-referendum deliberative process (Frey 1994), difficulties to translate political preferences into policy choices (Selb 2008); (iii) lower turnout or higher roll-off (Bowler, Donovan, and Happ 1992; Bowler and Donovan 1998; Reilly and Richey 2011), (iv) a stronger inclination to reject propositions Bowler and Donovan (1998), and (v) a stronger reliance on parliamentary recommendations (Stadelmann and Torgler 2013).

[5] Hessami (2016) analyzes the effect of proposition complexity on aggregate referendum outcomes using the number of subjects per proposition as a measure of complexity.

[6] For recent research and more institutional details on direct democracy in Switzerland we refer readers for instance to Funk and Gathmann (2013) and Hofer, Marti, and Bütler (2017).

We make three substantive contributions to the literature. Our first contribution is to develop a new measure for proposition complexity. Conceptually, we take a different approach than the existing literature by constructing a measure of the *underlying* complexity of ballot propositions.[7] The complexity of a proposition that voters are exposed to *prior* to a referendum is the variable of interest which we capture with our novel complexity measure based on an extensive data collection effort for 223 federal referendums in Switzerland over the 1981-2010 period. We use information provided in official pre-referendum booklets which the Swiss government is legally obliged to disseminate before each referendum since 1978 (Schweizer Bundesrat 1978). In constructing this measure, we follow the literature which regards complex propositions as those that are "lengthy (…) and technical" (Lupia 1994, 65). While the level of technicality is subjective and difficult to measure, the length of the description of propositions can be measured. Hence, we record the number of words in the information and debate section in the official booklets for each proposition.

Our second contribution is that to our knowledge we are the first to study the effect of complexity on individual voting behavior – or on voting behavior in direct legislation more generally – with an econometric specification which addresses sample selection. For any proposition voters face two decisions: (i) whether to participate in the referendum and (ii) conditional on participation whether to choose the status quo (reject the proposition) or the ballot measure (accept the proposition).[8] Since the second decision – the voting decision – is only observable for citizens that turn out, the classic *sample selection problem* arises (Heckman 1978, 1979). In our case, the selection bias is a participation bias.[9] Previous studies ignore sample selection and typically use aggregated data on the turnout rate or the share of yes- or no-votes as the dependent variable. We apply a structural approach based on two

[7] Existing measures (ballot complexity and proposition prolixity) conceptually target complexity at a more superficial level, i.e. the number of propositions on a ballot and the proposition content as stated on the ballot. Our measure of proposition complexity does not require citizens to read the booklets to be exposed to the complexity of a proposition. Instead, the length of the information text in the official booklet for a given proposition serves as a proxy of the underlying proposition complexity. In section VII.B of this paper, we provide evidence that effects do not differ between booklet readers and non-readers.

[8] In the context of our setting, the pre-reform situation is the status quo. Voters are asked to approve a policy change in the ballot measure. Hence, a rejection implies that the pre-reform situation will prevail and voters are more likely to prefer the pre-reform situation. In addition, note that the default in referendums is always the pre-reform situation. Therefore, a no-vote, which is conceptually a vote in favor of the status quo, is in effect a vote for the pre-reform situation. Finally, it should be noted that in Switzerland there is no participation quorum, i.e. regardless of turnout, the outcome of the referendum is binding.

[9] For propositions on immigration issues, Krishnakumar and Müller (2012) find a substantial participation bias of 17 ppts.

estimation equations (Heckman selection model). We solve the endogeneity problem by means of an exclusion restriction, i.e. we include a variable (the closeness of a referendum outcome) that influences the individual participation decision but which is arguably orthogonal to the voting decision. This allows us to portray more accurately how complexity affects individual voting behavior and to disentangle a direct effect of complexity on the voting decision and possible indirect effects via the participation decision.[10] This contribution should not be understood as a primarily technical matter, but also as an attempt to bring the empirical analysis closer to the structure of the underlying political-economic theory on the act of voting.

Our third contribution is to validate theoretical mechanisms proposed by the literature. We examine how the effect of complexity on individual voting behavior varies with voters' education level. Various contributions in the literature suggest that voters' decisions to participate in a referendum or to accept or reject a proposition depend on how educated they are. Matsusaka (1995) points out that citizens may be more likely to turn out when their issue-related knowledge and education level is high. Gerber and Lupia (1999) argue that voter characteristics that mitigate the uncertainty due to a proposition's content should reduce the status quo bias. Hence, better educated voters have to invest less effort to understand the content of a complex proposition and are less likely to rely on the status quo heuristic.

We provide evidence that voters are more likely to reject propositions when they are more complex. Increasing proposition complexity from the 10^{th} to the 90^{th} percentile would decrease the approval rate by 5.6 ppts. This decline is often decisive: an additional 12% of the propositions in our sample would be rejected. Two competing mechanisms determine the impact of complexity on the status quo bias in the vote outcome: a *direct effect* – confronted with higher complexity, voters tend to vote in favor of the status quo, – and an *indirect participation effect* – increasing complexity reduces the percentage of citizens biased towards the status quo (through vote abstentions) among the voters and therefore mitigates the bias. We contribute to the literature by estimating the magnitude of both effects. Neglecting the indirect participation effect would lead us to overestimate the effect of complexity on the referendum outcome by almost 50%. The sheer magnitude of the opposing indirect effect calls attention to the importance of treating voting behavior as an outcome of two sequential choices.

[10] Note that the Heckman estimations also allow us to estimate the size of the status quo bias in the entire population of a country (independently of the turnout decision) which would not be possible with ordinary one-equation estimations.

Voters with a lower ability to understand complex issues (i.e. with a lower education level) are on average more than 10% more likely than highly educated voters to abstain from voting and to reject propositions. This effect is twice as large for propositions at the 90th percentile compared to propositions at the 10th percentile of the complexity distribution in our sample. Our findings hence suggest that less educated citizens are more disinclined to turn out when propositions are highly complex. Lijphart (1997, 1) claims that such "unequal participation spells unequal influence" and thereby calls the legitimacy of referendums on complex issues into question. However, in line with the swing voter's curse theory, we find that citizens biased towards the status quo are more likely to abstain from voting which renders the vote outcome informationally superior since it mitigates the status quo bias in the vote outcome. By estimating the participation bias, we contribute a quantitative dimension to the discussion of the underlying trade-off between a *representative vote outcome* (Lijphart 1997) and an *informationally superior vote outcome* (Feddersen and Pesendorfer 1996).

Our objective complexity measure has the advantage that it is unrelated to individual characteristics and can thus be used to study the interaction between voters' characteristics (e.g. education level) and proposition complexity. In an additional extension we use an alternative measure of proposition complexity based on voters' perceptions of how difficult it was for them to form an opinion about the consequences of rejecting or accepting a proposition (subjective complexity). In this case, voters who find it rather difficult to form an opinion are 11.5 percentage points less likely to turn out and 4.7 percentage points more likely to reject a proposition.

We conduct five robustness tests that address potential concerns regarding our complexity measure, our identification strategy, and alternative mechanisms. First, we use data on a survey question that indicates whether a respondent has or has not used the official information booklet. We find that our estimation results for the effect of complexity on individual voting behavior do not differ between booklet users and nonusers. This indicates that our complexity measure indeed captures proposition complexity at a deeper level and that its validity does not require citizens to read the information booklets. Second, we show that our main results are not confounded by the fact that more complex propositions may be perceived by voters as more important. Third, we provide evidence that the government does not strategically manipulate proposition complexity to induce voters to vote in line with the government's recommendation. Fourth, using pre-poll data on the ex-ante closeness of referendum outcomes for a subset of our sample we obtain results similar to our baseline estimates. Fifth, we provide evidence that our baseline estimates are not contaminated by those propositions where survey bias – as identified by Funk (2016) – may be an issue.

The findings in this paper are interesting not only because they provide a comprehensive analysis of how complexity in direct legislation affects individual voting behavior, but also because direct democracy is becoming a more important tool for decision-making in various contexts (see e.g. the British referendum on EU membership or the Greek referendum on bailout packages). From a policy perspective, it is important to know how the complexity of policy issues influences individual voting behavior as well as aggregate referendum outcomes. Our results suggest that governments should invest more in general education to form politically mature citizens able to make informed decisions even if propositions are unusually complex.

II. Theoretical Considerations

A. *Proposition complexity and the participation decision*

Downs (1957) and Riker and Ordeshook (1968) introduce a theory of voting based on a rational trade-off between the costs of voting and the expected benefits. Matsusaka (1995) extends the traditional rational voter model by highlighting the role of limited information. Voters are portrayed as utility-maximizing consumers who receive higher payoffs from casting their vote when they are more confident of their vote choice. We extend this argument further by taking into account that the amount of information that a voter needs to reach a certain level of confidence depends on the complexity of the issue at stake. Matsusaka (1995) also states that information is meant to comprise issue-specific knowledge as well as general knowledge (education). Hence, when voters are more educated they are less inclined to abstain from voting.

In a similar vein, the swing voter’s curse theory argues that less informed voters may rationally prefer to abstain from voting even when they have a strict preference in favor or against a proposition and voting is costless (Feddersen and Pesendorfer 1996). Assuming that voters have homogeneous preferences but differ in their ability to identify the preferred choice, uninformed voters can only be pivotal if they vote differently than informed voters which is irrational since informed voters vote for the preferred option with certainty. Therefore, some uninformed voters may find it rational to vote against their prior beliefs as long as enough uninformed voters still vote in opposition to informed voters. If all uninformed voters vote against their prior beliefs, they may dominate informed voters and the inferior alternative will be elected. Hence, uninformed voters find it optimal to abstain from voting to maximize the probability that informed voters determine the vote outcome (Feddersen and Pesendorfer 1996).

The empirical implication of these two theories is that on average voters should be more reluctant to participate in a referendum with higher proposition complexity. This effect should be stronger for individuals with a low level of education.

B. *Proposition complexity and the voting decision*

Samuelson and Zeckhauser (1988) were among the first to draw wide attention to status quo bias in decision-making. They conducted a number of experiments showing that individuals disproportionately tend to stick with the status quo. The authors draw on a broad range of insights from economics, psychology, and decision theory to provide theoretical explanations for status quo bias. One of these explanations is the existence of uncertainty. The authors state that an early choice may have a substantial advantage over an alternative. From a consumer choice perspective, Samuelson and Zeckhauser (1988) argue that consumers remain loyal to a chosen brand as long as their utility from consuming this product is above a certain threshold. In the context of referendums this implies that as long as voters enjoy a minimum level of utility or do not significantly suffer from the status quo they may reject a proposition not even bothering to find out whether they might benefit from a policy change. The authors also argue that "the choice to undertake a decision analysis is itself a decision" (p. 35). If the costs of this analysis are high, voters may only conduct this analysis once, take a decision, and then defer to this choice in the future. If we apply these ideas to the context of direct democracy, one could argue that both the complexity of propositions as well as voters' education levels determine how large the cost of analysis is and whether a proposition is rejected.

Second, a key insight of Eichenberger and Serna (1996) is that individual errors in the assessment of expected benefits, even if random, have asymmetric effects on the referendum outcome due to the nature of the voting procedure. A proposition may benefit an average voter, i.e. the associated policy changes generally increase net-utility. For an individual voter, however, these policy changes may or may not be beneficial. The complexity of a proposition increases the variance of the expected benefits of a proposition for a given voter. When complexity is higher, a larger number of voters will believe that this proposition is very beneficial or very harmful for them, i.e. individual errors become larger. The larger number of voters who (wrongly) believe that the proposition is very beneficial for them is inconsequential for the referendum outcome. These voters would vote in favor of the proposition even if it were less complex. The larger number of voters who (wrongly) believe that the proposition is very harmful, however, has substantive consequences for the referendum outcome. Some of the voters who actually benefit from the proposition now underestimate its benefits and reject the proposition. This mechanism likely varies with the education level of voters. Random errors may be less relevant for more educated voters who are better equipped to process complex information or more specifically to gauge the personal costs and benefits of complex propositions. Related to this, Eichenberger and Serna (1996) state: “it is difficult to measure the complexity of an issue independently from the individuals’ human capital” (p.140).

Third, a behavioral literature emphasizes that heuristicsare primarily used in situations where people have to trade off the effort required in decision-making and the accuracy of one’s decision (Gigerenzer and Gaissmaier 2010). Therefore, better informed or better educated voters have to invest less effort and are less likely to rely on the status quo heuristic. Highly educated voters are therefore less likely than less educated voters to be biased towards the status quo even though they may also reject propositions because they are highly complex.

The empirical implication of these three theories is that on average voters should be more likely to reject propositions with increasing proposition complexity. This effect should be stronger for individuals with a low level of education.

III. Empirical Strategy

A. *Participation bias and endogeneity*

For any proposition that is at stake, voters face two decisions: (i) participation versus abstention and (ii) conditional on participation the status quo (reject the proposition) versus the ballot measure (accept the proposition). Since the vote decision is only observable for the subset of citizens that participate in the referendum, the classic *sample selection problem* may arise (Heckman 1978, 1979).

We hypothesize that the complexity of a proposition influences both the participation and the vote decision as summarized in the following binary choice models:

(1) $$Participate^{*} = \alpha\, Complexity + \boldsymbol{\beta}'\mathbf{x} + u, \quad Participate = 1 \text{ if } Participate^{*} > 0,$$
$$Participate = 0 \text{ otherwise.}$$

(2) $$Yes\text{-}vote^{*} = \alpha\, Complexity + \boldsymbol{\beta}'\mathbf{x} + \epsilon, \; Yes\text{-}vote = 1 \text{ if } Yes\text{-}vote^{*} > 0,$$
$$Yesvote = 0 \text{ otherwise.}$$

where vector **x** includes a set of control variables.[11] In equation (1), the complexity of the entire ballot matters for the participation decision (i.e. the sum of complexity across all propositions on a ballot), whereas in equation (2) the complexity of individual propositions matters for the voting decision.

According to Heckman (1979), self-selection of citizens into the voting sample can be interpreted as an omitted variable problem in the equation (2). Several solutions to this problem may come to mind. A first idea would be to restrict the sample to the voting population, i.e. to ignore the first equation and to only rely on the data of the subsample of citizens who actually voted. This approach ignores that complexity or any other variable may alter the decision to vote in favor of a proposition and may also change the composition of voters participating in the election. This would cause either an upward or downward bias in α depending on how voters and non-voters differ in their participation response to complexity. If an increase in complexity causes citizens, which otherwise would have voted against the proposition, to abstain from voting, then α (a measure of the magnitude of the status quo bias) would be downward biased in regression equation (2), overstating the extent of the status quo bias.[12] A second potential solution is to control for participation in the vote decision equation and to use an instrument for the potentially endogenous participation decision. However, an IV approach is not feasible since the voting decision is only observable for citizens who participate in a referendum. A third potential solution for the omitted variable problem in equation (2) is to control for all characteristics of the participation decision by adding additional variables to equation (2). Even after controlling for all observable characteristics, the selection process might still be driven by unobservable factors such as the ability to cope with complex propositions or preference parameters like risk aversion (Samuelson and Zeckhauser 1988).

Our solution to the selection problem is to use a Heckman selection model which is identified by means of an exclusion restriction, i.e. we include a variable in equation (1) that influences the participation decision but which is arguably orthogonal to the vote decision. In doing so, we can estimate the magnitude of the selection bias – which is synonymous to the participation bias – and correct our estimates for the effect of complexity on the individual voting decision for this bias.

[11] For notational convenience, we use the same Greek letters indicating the coefficients to be estimated in equation (1) and (2). However, they can represent different estimates in each equation. We stick to this convention throughout the paper.

[12] In equation (2), a status quo bias corresponds to a negative estimate for α.

B. *Exclusion restriction and Heckman selection approach*

Without an exclusion restriction in equation (1), identification would solely rely on the bivariate normality assumption for the functional form of the error terms. Wooldridge (2010) shows that identification based on this assumption alone can be misleading and produce spurious results. In our setting, a valid exclusion restriction requires a variable that influences participation but that has no direct effect on the voting decision.

Based on a sizable literature on the relationship between the closeness of elections and turnout, we include the ex-post closeness of a referendum outcome as a valid exclusion restriction.[13] A positive correlation between the closeness of an election and the individual likelihood of casting one's vote is firmly grounded on various theoretical arguments. The seminal literature asserts that the benefit of voting increases with the probability of casting the decisive vote (Downs 1957; Riker and Ordeshook 1968). This probability is higher when an election is contested. This has been denoted in the literature as the *Downsian Closeness Hypothesis* (Matsusaka and Palda 1993). A second explanation for this positive relationship relies on more mobilization efforts (lowering participation costs) by stakeholders in contested elections which affects the individual probability to participate in the referendum (Cox and Munger 1989; Denver and Hands 1974; Key and Heard 1984). On the other hand, there is no reason why a close election would make it more likely that voters support or reject propositions.[14]

Our Heckman selection approach can be described by the following two equations:

(3) $$Participate^{*} = \alpha\, Complexity + \boldsymbol{\beta}'\mathbf{x} + \boldsymbol{\delta}'\mathbf{z} + u,\ Participate = 1 \text{ if } Participate^{*} > 0,$$

$$Participate^{*} = \alpha\, Complexity + \boldsymbol{\beta}'\mathbf{x} + \delta z + u,\ Participate = 0 \text{ otherwise},$$

[13] An alternative exclusion restriction that was often suggested to us is rainfall or bad weather conditions in general. We have run additional estimations (available upon request) using various weather indicators (rainfall during different time intervals during the day, snowfall and average temperatures). The results indicate that there is hardly any effect of bad weather on the participation decision. This is not surprising given that it is ex ante not clear whether bad weather leads people to abstain from voting (disinclination to go outside) or makes citizens more likely to participate (no other plans in bad weather). In addition, there is recent evidence that bad weather is likely not orthogonal to the voting decision and is therefore an invalid exclusion restriction (Meier, Schmid, and Stutzer 2016).

[14] There is no correlation between the individual vote decision and the closeness of elections because the absolute number of valid votes for each proposition in our sample period (1981-2010) is on average 2 million. Therefore, the individual voting decision has a negligible influence on the share of yes-votes. Another thought experiment illustrating the validity of our exclusion restriction is the following. Imagine a new variable for the share of yes-votes is constructed where we exclude the vote of one individual. The value of the adjusted share of yes-votes would change only to a very small degree, which would not affect our estimation results.

(4) $$Yes\text{-}vote^* = \alpha\, Complexity + \boldsymbol{\beta}'\mathbf{x} + \epsilon,\ Yes\text{-}vote = 1 \text{ if } Yes\text{-}vote^* > 0,$$

$$\underline{Yesvote^* = \alpha\, Complexity + \boldsymbol{\beta}'\mathbf{x} + \epsilon,}\ Yesvote = 0 \text{ otherwise,}$$

$$\text{with } \begin{pmatrix} u \\ \epsilon \end{pmatrix} | \mathbf{x}, \mathbf{z} \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix})))))9999 \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right],$$

where $\mathbf{z}$ captures the exclusion restriction.[15] The error terms u and ϵ are assumed to be distributed bivariate normal with $(u, \epsilon) \sim\ bivariate\ normal[0, 0, 1, 1, \rho]$, where ρ denotes the correlation between the error terms. By estimating ρ within the Heckman selection model, we are able to control for unobserved factors influencing both the turnout and voting decision. An estimate of ρ different from zero would point towards the presence of a selection bias justifying the Heckman selection model as preferred estimation strategy over the simple probit estimates based on equations (1) and (2). As in section III.A, in the participation equation complexity is aggregated at the ballot level, whereas in the voting equation complexity is measured at the proposition level. Vector **x** includes a number of control variables for standard voter characteristics: female dummy, age, education level, knowledge about the proposition, married dummy, Protestant dummy, employed dummy.[16] We additionally include dummies for the canton in which the respondent is living, the year in which a referendum is held, the type of referendum, and the policy area in which a proposition falls.[17] The inclusion of these controls is a straightforward way to control for a selection bias in equation (4), if selection is only driven by observables. However, it is unlikely that voters' preferences as well as their capacity to understand complex proposition can be fully accounted for by the inclusion of standard socio-economic variables. To deal with unobserved factors that drive self-selection into the voting population, we conduct Heckman estimations.

We use the log of the number of words for the information text of a proposition as the complexity measure (more details follow in section IV.B). We expect that the same absolute increase in the number of words of the information text has a stronger effect on voting behavior for instance for an increase from 100 to 200 words than for an increase from 1000 to

[15] We include the closeness of the referendum outcome and its square since we expect an inversely U-shaped relationship.

[16] Brunner, Ross, and Washington (2011) show that economic conditions shape preferences on direct-democratic legislation. In our estimations, this would at least be partially captured by the employed dummy. Funk and Gathmann (2015) provide evidence that female voters make different choices on direct-democratic propositions in Switzerland than male voters.

[17] Status quo bias and the willingness to participate in referendums may differ across cantons for cultural reasons. Year dummies allow capture common shocks in specific time periods as well as trends in participation and vote decisions. Policy area and referendum type dummies reduce the variation in proposition complexity due to these observable aspects.

1100 words, i.e. relative rather than absolute differences are relevant for voters. All hypothesis tests are based on standard errors that are clustered at the ballot level. As suggested by Freedman and Sekhon (2010), we solve the two-equation model in equations (3) and (4) using full information maximum likelihood estimation instead of the two-step procedure originally introduced by (Heckman 1978, 1979). [18,19]

IV. Data Description

A. *Post-referendum survey data*

We use data from standardized and representative polls conducted after each national referendum in Switzerland since 1981. The GfS Research Institute in Berne conducts these surveys on behalf of the Institutes of Political Science at the Universities of Berne, Geneva, and Zurich (FORS - Swiss foundation for research in social sciences 2012).[20] A random sample of 700 to 1000 eligible voters is selected from the Swiss telephone book and surveyed within two weeks after the elections.[21]

The VOX survey asks citizens about their participation in each referendum as well as their individual voting decision. These are our dependent variables in equations (3) and (4). The respondent is also asked about his knowledge about the proposition, the kind of media consulted prior to the referendum, the perceived importance of the vote, and various personal characteristics (age, gender, education, marital status, profession, etc.). Summary statistics for all variables at the respondent level are reported in Table A.1 in the online appendix.

B. *Official information booklets and complexity measure*

Since 1978, the Swiss government is obliged by law to mail a written information booklet before each national referendum to all eligible voters (Schweizer Bundesrat 1978). The office in charge of writing the information booklet ("Bundeskanzlei") has to follow strict

[18] The log-likelihood which has to be estimated is given by: $\ln L\left(\alpha_\pi, \alpha_v, \boldsymbol{\beta_\pi}, \boldsymbol{\beta_v}, \boldsymbol{\delta}; v, \pi, c_\pi, c_v, \mathbf{x}, \mathbf{z}\begin{matrix} 1 & \rho \\ \rho & 1 \end{matrix}\right) = \sum_{\pi,v=1} \ln \Phi_2(\alpha_\pi c_\pi + \boldsymbol{\beta}'_{\boldsymbol{\pi}}\mathbf{x} + \boldsymbol{\delta}'\mathbf{z},\ \alpha_v c_v + \boldsymbol{\beta}'_{\mathbf{v}}\mathbf{x},\ \rho) + \sum_{\pi=1,v=0} \ln \Phi_2(\alpha_\pi c_\pi + \boldsymbol{\beta}'_{\boldsymbol{\pi}}\mathbf{x} + \boldsymbol{\delta}'\mathbf{z},\ -\alpha_v c_v - \boldsymbol{\beta}'_{\mathbf{v}}\mathbf{x},\ -\rho) + \sum_{\pi=0} \Phi(-\alpha_\pi c_\pi - \boldsymbol{\beta}'_{\boldsymbol{\pi}}\mathbf{x} - \boldsymbol{\delta}'\mathbf{z})$. The variables *Participate* and *Yes-vote* are denoted by π and v. Parameters corresponding to the participation equation (3) are denoted by a subscript π, whereas parameters denoted by v correspond to the voting equation (4). Φ_2 represents the bivariate normal cdf. Φ denotes the standard normal cdf.

[19] The first study that estimated two probit equations in a Heckman selection framework is Van de Ven and van Praag (1981).

[20] The data is available free of charge at http://forscenter.ch/en/data-and-research-information-services/2221-2/obtain-data/

[21] The interviewer calls, introduces himself and asks whether there is an eligible voter in the household. If there are several eligible voters in a household, the one who has his birthday on the earliest day in the year is interviewed.

legal rules regarding the content of the information material. The booklets are required to be short, objective, transparent, and in line with the principle of proportionality (Bundesgericht 2008). It is explicitly forbidden by law to influence the decision-making process of voters towards accepting the proposition.[22]

The booklets on average have a total size of around 50 pages. Each proposition has a separate chapter in the booklet which consists of four sections: a short summary, a detailed information section, a debate section comparing arguments against and in favor of the proposition, and a legal section in which those parts of the law are published that would change if the referendum is successful.

We construct a novel proposition complexity measure based on information provided in the official booklets. We use standard office software that transfers the booklets into a machine-readable format allowing us to count the number of words in the information and debate section for each proposition. Highly complex propositions are associated with a longer description in the booklet. The strict legal framework requires a short and balanced booklet text and prohibits that the government agency in charge of writing the information booklets influences voters by exaggerating the views of the government. We therefore argue that the length of the information text is determined solely by the necessity to provide longer descriptions of more complex propositions to ensure that the content of the information booklets complies with legal requirements.[23]

Several mechanisms may lead to a positive relationship between the complexity of a proposition and the number of words used in its information text. First, if the content of the proposition itself is difficult to understand, one would expect a more extensive description to make the topic of the proposition accessible to average citizens. Second, a complex proposition that implies various individual policy measures in a certain policy area requires a description of each policy measure and therefore requires a longer description in the official booklet.

[22] In a robustness test (see section VII.D), we find no empirical evidence for a correlation between the government recommendation and our text-based complexity measure.

[23] An alternative measure for complexity could be constructed based on the legal text that would change if a referendum is successful. Each proposition gives rise to a change to the constitution or to existing laws. Huber and Shipan (2002), however, argue that the detailed language and the resulting length of the legal text might be driven by politicians' incentives to delegate policy making to other policymaking authorities such as bureaucrats.

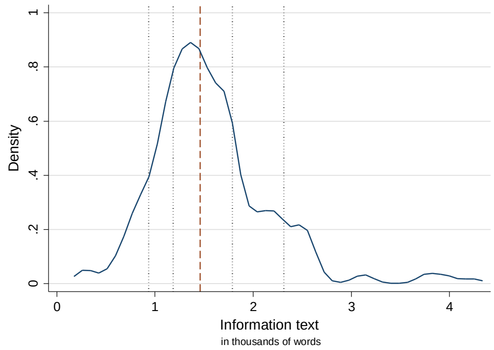


FIGURE 1. DISTRIBUTION OF THE PROPOSITION COMPLEXITY MEASURE

Notes: This figure depicts the variation in our proposition complexity measure using a Gaussian kernel density plot with a kernel bandwidth of 100 words. The dashed red line represents the median of the complexity measure. The dotted grey lines correspond to the 10th, 25th, 75th, and 90th percentiles, respectively. For better readability, the information text axis is restricted to values below the 99th percentile.

Figure 1 illustrates the large variation in proposition complexity in our sample of 223 federal referendums in Switzerland between 1981 and 2010. The median length of the information text is around 1,500 words; roughly 80 percent of the observations lie between 900 and 2,300 words.

V. Estimation Results

A. *Probit estimations for the participation and vote decision*

In this section, we report estimation results that we obtain when ignoring participation bias. We estimate two separate estimation equations (see equations (1) and (2)) using the probit estimator. The benefit of this exercise is to obtain a benchmark based on a naïve empirical strategy which allows us to assess how the results change when we do account for participation bias (see section V.B).

TABLE 1—PROBIT ESTIMATION RESULTS: COMPLEXITY AND THE INDIVIDUAL PARTICIPATION AND VOTE DECISION

Avg. Marginal Effects reported	Dep. Var.: Participation				Dep. Var.: Yes-Vote			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Complexity (ballot)	-0.098***	-0.104***	-0.115***	-0.122***				
	(0.036)	(0.036)	(0.043)	(0.036)				
Complexity (proposition)					-0.085**	-0.081**	-0.093***	-0.105***
					(0.040)	(0.035)	(0.034)	(0.030)
Rural	-0.003	-0.010**	-0.008*	-0.008	-0.040***	-0.035***	-0.036***	-0.037***
	(0.005)	(0.004)	(0.005)	(0.005)	(0.007)	(0.007)	(0.007)	(0.007)
Female	-0.003	-0.004	-0.004	-0.002	0.020***	0.023***	0.024***	0.024***
	(0.006)	(0.006)	(0.006)	(0.006)	(0.008)	(0.007)	(0.006)	(0.006)
Age	0.005***	0.005***	0.005***	0.005***	-0.001**	-0.000*	-0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Education	0.044***	0.044***	0.036***	0.035***	0.016***	0.017***	0.018***	0.018***
	(0.003)	(0.003)	(0.002)	(0.002)	(0.005)	(0.004)	(0.003)	(0.003)
Proposition Knowledge	0.056***	0.056***	0.065***	0.068***	0.027**	0.031***	0.028***	0.028***
	(0.005)	(0.005)	(0.005)	(0.004)	(0.012)	(0.008)	(0.007)	(0.006)
Married	0.071***	0.069***	0.073***	0.072***	-0.012*	-0.012**	-0.013***	-0.013***
	(0.007)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.005)	(0.005)
Protestant	0.017***	0.011*	0.021***	0.020***	-0.006	-0.009*	-0.007	-0.007
	(0.005)	(0.006)	(0.005)	(0.005)	(0.007)	(0.005)	(0.004)	(0.005)
Employed	-0.002	-0.006	-0.005	-0.005	0.003	0.001	-0.001	-0.001
	(0.006)	(0.006)	(0.006)	(0.005)	(0.006)	(0.006)	(0.006)	(0.006)
Canton dummies	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Referendum type dummies	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Year dummies	No	No	Yes	Yes	No	No	Yes	Yes
Policy area dummies	No	No	No	Yes	No	No	No	Yes
Pseudo R^2	0.140	0.149	0.164	0.172	0.009	0.058	0.070	0.087
Observations	204818	204818	204818	204818	107420	107420	107420	107420

Notes: The table establishes the significantly negative effect of complexity on the likelihood that a voter participates in a referendum or votes vote in favor of a proposition. *Complexity* represents the log of the number of words used in the information text in the official booklets per proposition (voting equation) or aggregated at the ballot level (participation equation). Average marginal effects based on probit regressions are reported in all specifications. The marginal effect of age is based on age and its squared term. Heteroskedasticity-robust standard errors clustered at the ballot level are reported in parentheses. *** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level.

Table 1 collects the regression results for the participation (Models (1) to (4)) and voting decision (Models (5) to (8)).[24] We report average marginal effects instead of probit coefficients. Note that the sample size for the estimation of the voting decision is about half as large ($N = 107420$) as the sample used for the estimation of the participation decision ($N = 204818$) since we observe the voting decision only for citizens who participated in the referendum.

The results for Models (1) to (4) suggest that complexity is negatively and significantly related with the probability of participating in a referendum.[25] The magnitude of

[24] Summary statistics for the variables used in the regressions are provided in Table A.1 in the online appendix.

[25] The high significance levels in our statistical tests are not driven by the large sample size. Our objective complexity measure varies only at the proposition or ballot level. Therefore, we correct our standard errors by clustering at the ballot level, allowing observations within a ballot to be correlated. The power of our statistical tests is therefore determined by the number of independent observations (ballots) in our estimation sample. With respect to the relatively low number of 74 ballots, high significance levels cannot be attributed to the size of the entire sample, but rather to a large quantitative effect

the negative effect of complexity on participation is considerable. Based on our estimates in Model (4), when complexity increases by one standard deviation (i.e. the information text is 3992 words longer at the ballot level), citizens on average have a 6.4 ppts lower probability to participate in the referendum.[26] These results are robust to the inclusion of fixed effects for cantons, referendum type, years, and policy areas. The results also show that more educated, more politically interested and more knowledgeable voters are significantly more likely to participate in a referendum. A reasonable explanation is that these voters have to invest fewer resources to estimate the consequences of their voting decision. The results for Models (5) to (8) show that proposition complexity has a significantly negative effect on the probability of voting in favor of a proposition. Based on our estimates in Model (8), when complexity increases by one standard deviation (i.e. the information text is 895 words longer at the proposition level), citizens are on average 4.6 percentage points less likely to vote in favor of a proposition. The inclusion of fixed effects slightly increases the precision (lower standard errors) and the size of the estimated marginal effect.

In Figure 2, we plot the results for the most complete models (Models (4) and (8)) to discuss in more detail the size of our estimates.[27] The shaded areas indicate 95 percent confidence intervals. The vertical dotted lines illustrate the distribution of proposition complexity in our sample by indicating the 10th, 25th, 50th, 75th, and 90th percentiles.

(as we will illustrate later in this section) and maybe to relatively low noise in our estimates for the effect of complexity on voting behavior.

[26] This value is calculated based on a centralized change by one standard deviation in complexity (log of word count).

[27] We follow McCloskey and Ziliak (1996) and Ziliak and McCloskey (2008), suggesting that the size of the estimate is at least as important as the statistical significance. The following graphical representation of the nonlinear relationship relies on comments in Wooldridge (2004) and Greene (2010).

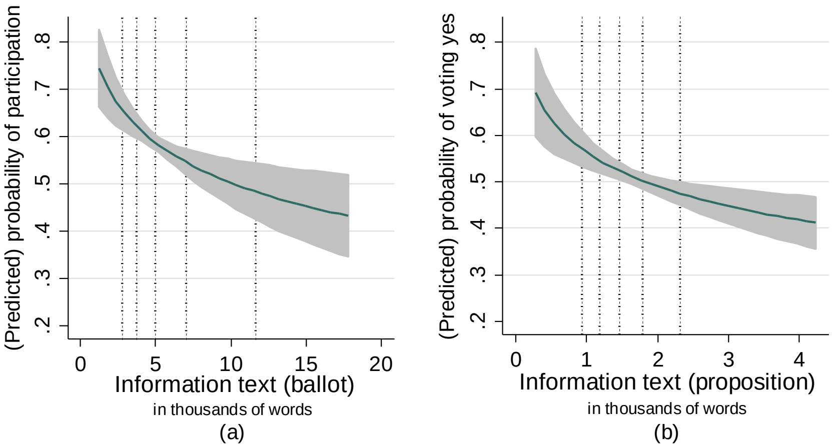


FIGURE 2. EFFECT OF COMPLEXITY ON THE INDIVIDUAL PARTICIPATION AND VOTING DECISION

Notes: This figure depicts the statistically significant negative effect of complexity on the probability that a voter participates in a referendum (subfigure (a)) and votes in favor of a proposition (subfigure (b)). In both cases, we plot average predicted probabilities against complexity. The estimates in subfigures (a), and (b) are calculated based on the estimation results for Models (4) and (8) in Table 1. The shaded area in subfigures (a) and (b) represents the 95 percent confidence interval band of the predicted probabilities. The dotted vertical lines correspond to the 10^{th}, 25^{th}, 50^{th}, 75^{th}, and 90^{th} percentiles of complexity, respectively. For better readability, the information text axis is restricted to values below the 99^{th} percentile is plotted.

According to Figure 2, the predicted participation rate for a low-complexity ballot (at 90^{th} percentile) is around 65% and falls below 50% for a high-complexity ballot (at 10^{th} percentile). The predicted probability of voting in favor of a proposition drops from 58% to 49% when comparing a low-complexity proposition with a high-complexity proposition. Our estimation results based on a naïve empirical strategy point toward substantial status quo bias due to complexity. In the next section, we will apply a more suitable empirical strategy to account for participation bias to obtain unbiased estimation results.

B. *Heckman estimations to identify and correct for participation bias*

As described in section III.B, we include the variables *closeness* and *closeness squared* in our estimations to implement the exclusion restriction.[28] To test the validity of this approach, we regress the participation dummy on our measure for the voter's expectation of a narrow voting decision using a probit estimator. Table 2 reports the regression results. We

[28] Instead of relying on the quadratic specification, we could use a linear measure for the absolute distance of the share of yes-votes from the 0% threshold (see Figure 3). Our approach, however, is preferable. It allows the effect to be nonlinear, thereby we do not impose a hump-shaped relationship with a peak around 0%, but it is a result of the estimation. We therefore not only test whether closeness turns out to be a statistically significant regressor, but also whether the implied quadratic functional form is indeed hump-shaped (negative coefficient for *closeness squared*).

find a statistically significant hump-shaped relationship (see Figure 3). The highest participation rate is indeed associated with a close outcome.[29]

TABLE 2—CLOSENESS OF THE REFERENDUM OUTCOME AND PARTICIPATION

Probit coefficients reported	Dep. Var.: Participation			
	(1)	(2)	(3)	(4)
Complexity (ballot)	-0.322***	-0.326***	-0.404***	-0.412***
	(0.110)	(0.111)	(0.124)	(0.109)
Exclusion Restriction				
Closeness	-0.355**	-0.303	-0.439***	-0.349**
	(0.153)	(0.204)	(0.169)	(0.167)
Closeness squared	-2.076***	-1.980***	-2.223***	-2.058***
	(0.579)	(0.617)	(0.678)	(0.668)
p-value for joint significance of linear and quadratic terms in:				
Closeness	0.000	0.003	0.001	0.003
Canton dummies	No	Yes	Yes	Yes
Referendum type dummies	No	Yes	Yes	Yes
Year dummies	No	No	Yes	Yes
Policy area dummies	No	No	No	Yes
Pseudo R^2	0.140	0.149	0.164	0.172
Observations	204818	204818	204818	204818

Notes: The table illustrates the significant hump-shaped effect of closeness on the probability to participate in a referendum. Regression coefficients based on probit regressions are reported in all specifications. All equations are estimated including the same individual controls as in Table 1. Heteroskedasticity-robust standard errors clustered at the ballot level are reported in parentheses. *** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level.

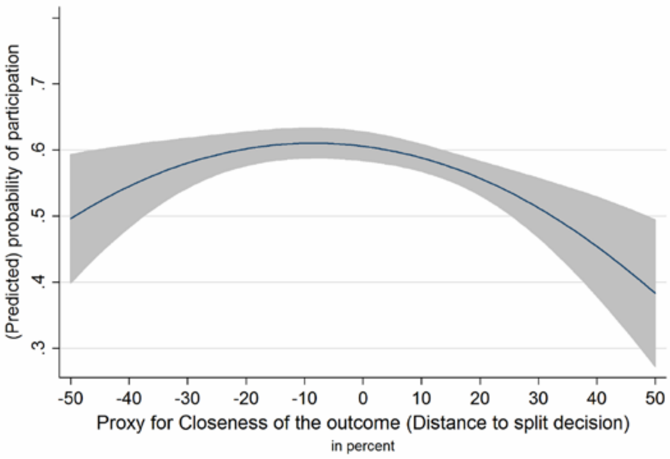


FIGURE 3. (EXPECTED) CLOSENESS AND THE PARTICIPATION DECISION

Notes: This figure illustrates the hump-shaped relationship between the closeness of a referendum outcome and the probability that a citizen participates in a referendum. We plot the average predicted probability against the share of yes-votes. The estimate is calculated based on the results of the probit regression presented in Model (4) in Table 2.The figures based on Models (1) to (3) are qualitatively and quantitatively similar. The shaded area represents the 95 percent confidence interval band of the predicted probability of participation.

Table 3 presents the estimated coefficients for the Heckman selection model.[30] The coefficients for the complexity measures at the proposition and ballot level have the expected

[29] A second condition for the validity of our exclusion restriction is that the expectation of a narrow outcome has no direct effect on the voting decision. This does not exclude the possibility of an indirect effect via the decision to participate, which does not violate the assumptions regarding to a valid exclusion restriction in the Heckman selection model.

negative sign and are significant at the 1 percent level. The estimated coefficient ρ measures the correlation between the error terms of the participation and the voting equation and can be interpreted as a measure of unobserved factors affecting both the participation and outcome decision. The estimate for ρ is positive and statistically significant in all specifications. [31] This implies that unobserved factors affect the probability to participate and the probability to vote in favour of the proposition in the same direction.[32] Hence, the indirect effect of complexity on the voting decision through the participation decision is positive. Higher complexity increases the probability that citizens (who otherwise would have voted against the proposition) are overwhelmed by complexity and therefore abstain from voting.

[30] Note that the number of observations is smaller in Table 3 than in Table 1 (191669 instead of 204818). In the two-equation Heckman model any missing observation in the voting equation also leads to a missing observation in the participation equation.

[31] Our identification strategy rests on the exclusion restriction Z which is included in equation (3) but not in equation (4). Equations (3) and (4) also differ in terms of the complexity measure that is used. In the participation equation, complexity is aggregated at the ballot level, while in the voting equation complexity is captured at the proposition level. We have run additional Heckman regressions where the exclusion restriction Z is neither included in equation (3) nor in equation (4). The results (available upon request) show that the identification of ρ is exclusively driven by the instrument Z.

[32] In all specifications the correlation is decreasing as we include further controls and fixed effects. This nicely illustrates the interplay between observables and unobservables in the model. Controlling for observable determinants decreases the role that unobservables play in determining the participation and voting decision. However, even in Model (4) in Table 3, where we control for differences in cantons, type of referendum, years, and policy areas (as well as socio-economic factors, which are included in all specifications in Table 3), the correlation coefficient ρ is still quantitatively large and precisely measured.

TABLE 3—RESULTS FOR HECKMAN SELECTION MODELS: COMPLEXITY AND THE INDIVIDUAL PARTICIPATION AND VOTE DECISION

Heckman coefficients reported	Heckman (1)		Heckman (2)		Heckman (3)		Heckman (4)	
	(1-1)	(1-2)	(2-1)	(2-2)	(3-1)	(3-2)	(4-1)	(4-2)
Dep. Variable:	Participation	Yes-Vote	Participation	Yes-Vote	Participation	Yes-Vote	Participation	Yes-Vote
Complexity (proposition)		-0.266***		-0.282***		-0.331***		-0.341***
		(0.054)		(0.082)		(0.082)		(0.079)
Complexity (ballot)	-0.357***		-0.384***		-0.479***		-0.502***	
	(0.106)		(0.121)		(0.134)		(0.119)	
Exclusion Restriction								
Closeness	-1.167***		-0.840**		-0.948***		-0.721***	
	(0.130)		(0.340)		(0.260)		(0.241)	
Closeness squared	-1.918***		-2.026***		-2.208***		-2.149***	
	(0.506)		(0.646)		(0.763)		(0.755)	
Unobserved Factors								
rho	0.892		0.555		0.517		0.395	
Wald test (*p-value*)	0.000		0.033		0.005		0.001	
Canton dummies	No		Yes		Yes		Yes	
Referendum type dummies	No		Yes		Yes		Yes	
Year dummies	No		No		Yes		Yes	
Policy area dummies	No		No		No		Yes	
Observations	191669		191669		191669		191669	

Notes: The table provides the estimated coefficients of the Heckman selection model and establishes the negative and significant effect of complexity on voter's probability to participate and vote in favor of a proposition. The table also reports the correlation ρ between the error terms of both equations, as well as the corresponding p-values. *Complexity* represents the log of the number of words used in the information text in the official booklets per proposition (voting equation) or aggregated at the ballot level (participation equation). All equations are estimated including the same individual controls as in Table 1. Heteroskedasticity-robust standard errors clustered at the ballot level are reported in parentheses. *** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level.

In the following, we examine in more detail the nature of the participation bias. The participation bias arises due to observed and unobserved factors which affect both decisions. Table 4 reports estimates for the resulting participation bias based on the regression results in column (4) in Table 3. Whereas on average only 46.8% of those citizens who are against the proposition participated in the referendums, 67.8% of the citizens who are in favour of the proposition participated in the referendums. The resulting participation bias in the average referendum outcome equals roughly 11 ppts.

TABLE 4—PARTICIPATION BIAS OF THE VOTING OUTCOME

Voting Preferences and Behavior		
Voting Yes	$Pr(v=1 \mid \pi=1)$	0.525
Preferring Yes	$Pr(v=1)$	0.417
Participation Bias	$Pr(v=1 \mid \pi=1) - Pr(v=1)$	0.108
Participation Behavior		
Participation	$Pr(\pi=1)$	0.555
Participation of those preferring Yes	$Pr(\pi=1 \mid v=1)$	0.678
Participation of those preferring No	$Pr(\pi=1 \mid v=0)$	0.468

Notes: The table establishes the resulting participation bias of 10.8 percentage points and illustrates that potential yes-voters are more likely to participate than potential no-voters. The estimates are based on the model estimates in column (4) of Table 3.

Since coefficients in nonlinear models (especially when these coefficients are associated with variables appearing in both the selection and the outcome equation of a Heckman selection model) are difficult to interpret, we report average partial effects for our main variables of interest in Table 5. The indirect effect of complexity on the vote outcome

$(v = 1|\pi = 1)$ – that can be identified with the Heckman approach – is positive. An increase of one standard deviation in complexity increases the average probability of voting in favor of a proposition by 2.3 percentage points (fourth column in Table 5). Higher complexity reduces the turnout rate of potential no-voters more strongly than for potential yes-voters. However, the indirect effect is quantitatively not large enough to offset the negative direct effect of complexity on the voting decision (-5.4 percentage points).

TABLE 5—PARTIAL EFFECTS OF HECKMAN MODELS

APE of ± 0.5 SD	Probit	Probit	Heckman		
	π=1	v=1|π=1	π=1	v=1|π=1	v=1
Complexity (proposition)		-0.046*** (0.013)		-0.054*** (0.013)	-0.051*** (0.011)
Complexity (ballot)	-0.082*** (0.020)		-0.082*** (0.019)	0.023*** (0.007)	

Notes: The table summarizes the average partial effects of a change of one standard deviation (centered, ± 0.5 SD) in each complexity measure on participation and voting behavior based on the single equation models (column (4) in Table 2, and column (8) in Table 1) and the Heckman selection model (column (4-1) and (4-2) in Table 3). The Heckman model allows for an indirect effect of the variable Complexity (ballot) on the vote outcome via altering the participation decision. This indirect effect increases the probability of voting in favor of a proposition by 2.3 percentage points if complexity changes by one SD (centered). Even though quantitatively important the indirect effect is outweighed by the negative direct effect of complexity on the probability of voting in favor of a proposition of -5.4 percentage points. All results are based on regression estimates using the estimation sample used in the Heckman regressions (n=191669). *** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level.

The last column in Table 5 provides the estimate of the effect of proposition complexity on the voting behavior of the entire electorate ($\Pr(v = 1)$), including voters and non-voters. This estimate can be interpreted as the effect of complexity on the preference in favor of an approval of a proposition of the country's population.

In Table 4, the difference between the average probability of voting in favour of a proposition ($\Pr(v = 1|\pi = 1) = 52.5\%$) and the average probability of preferring an approval of a proposition ($\Pr(v = 1) = 41.7\%$) becomes evident. An increase of proposition complexity by one standard deviation reduces the average preference in the population for an approval of the proposition by 5.1 percentage points (see last column in Table 5). This effect can be interpreted as the status quo bias in the entire population of a country, independently of the turnout decision.[33]

To investigate the consequences of neglecting selection effects in the estimation approach with respect to the prediction of the vote outcome, we compare the Heckman results with the results from the naïve single equation probit specifications as described in equations (1) and (2). Based on the estimates in Table 5, one might be tempted to accept the probit estimate for the average effect of complexity on voting behavior (-4.6 ppts) as a reasonable

[33] With a single-equation regression model (such as the probit model in equation (2)), which focuses on the behavior of voters, these results cannot be obtained.

approximation for the Heckman estimates consisting of both direct (-5.4 ppts) and indirect (2.3 ppts) effects of complexity. Yet, the direction of the bias of the probit estimates is systematically related to the complexity of the ballot.

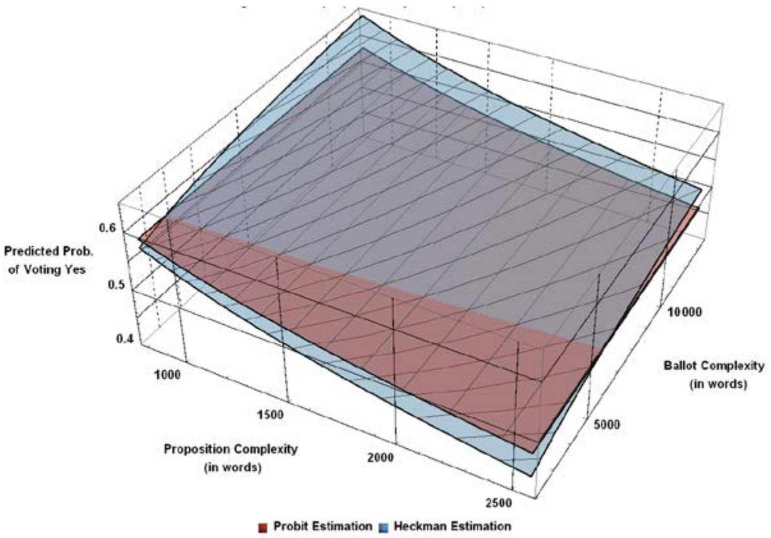


FIGURE 4. PROBIT VS. HECKMAN ESTIMATION RESULTS: PREDICTED PROBABILTIES OF VOTING IN FAVOR OF A PROPOSITION

Notes: This figure compares the results for the naïve probit estimations in section V.A with the Heckman estimation results in section V.B for the effect of complexity on individual voting behavior. We plot the average predicted probability of voting in favor of a proposition against the complexity of the proposition (direct effect) and the complexity of the ballot (indirect effect). The probit estimation (equation (2)) neglects the indirect effect. Therefore, the probit estimations are independent of ballot complexity. The estimates in the above figure are based on the estimates for Model (4) in Table 3. The dotted vertical lines correspond to the 10^{th}, 25^{th}, 50^{th}, 75^{th}, and 90^{th} percentiles of the respective complexity measure, respectively. For better readability, the information text axes are restricted to values between the 5^{th} and 95^{th} percentile.

Figure 4 compares the average predicted probability to accept a proposition against ballot and proposition complexity for both the probit (red surface) and the Heckman model (blue surface). The fact that the two surfaces intersect shows that depending on the level of complexity, the probit estimator tends to either over- or underpredict the effect of complexity on voting yes. Probit estimates neglect the indirect effect of ballot complexity on the vote outcome. Therefore, the probit model underestimates the probability of voting in favor of a proposition in cases in which complexity of the corresponding ballot and therefore the size of the opposing a positive indirect effect is very high. Vice versa, the probit model overestimates the probability of voting in favor of a proposition when ballot complexity is low. The difference in the predictions between the probit and Heckman estimation results is quantitatively substantial. When complexity is low the probit estimator sometimes predicts a share of yes-votes larger than 50% and the acceptance of the proposition, while the Heckman estimator predicts a share below 50% and the rejection of the proposition. The opposite occurs more likely for highly complex ballots for which the probit model predicts a too low share of yes-votes.

C. *Simulations of the policy impact*

We conclude the presentation of the Heckman estimation results with a simulation exercise. We investigate to what extent proposition complexity may decisively influence a referendum outcome, i.e. tilt the aggregate outcome from approval to rejection and vice versa. We fix complexity at the ballot and proposition level at the 10th percentile and calculate individual predictions for each of the 191,669 observations in the sample. Afterwards, we repeat this exercise for the 90th percentile.[34] We refer to these as *low* and *high* complexity scenarios.

In line with the results in Table 5, we find that the direct effect of proposition complexity (ignoring the participation effect) leads to a reduction in the probability of voting in favor of a proposition by 11.8 ppts. However, the participation effect partially offsets the decline in approval. The predicted participation rate decreases by 21.6 ppts. Since the participation rate of citizens who would vote against the proposition declines disproportionately, the participation effect of complexity (via the participation effect) on the approval rate leads to an increase of 6.1 ppts. In total, an increase in ballot and proposition complexity from the 10th percentile to the 90th percentile causes the approval rate to decline by 5.6 ppts. Relying on a simple probit estimation (for the subsample of voters, ignoring the participation effect) would result in a predicted decline in the approval rate by 9.4 ppts.

In the final step, we average the individual predictions across propositions to obtain a collapsed data set containing the predicted approval rate for 223 propositions. Is complexity likely to alter the referendum outcome? We investigate in how many cases complexity causes the approval rate to decline below 50%. Figure 5 plots the cumulative distribution of the 223 propositions with respect to their (predicted) approval rate in the case of the low (red line) and high (blue line) complexity scenario. Subfigure (a) illustrates the joint complexity effect based on the Heckman approach, while subfigure (b) only illustrates the direct effect, ignoring the indirect participation channel.

[34] For the proposition complexity measure, the increase in complexity from the 10th to the 90th percentile represents an increase from 937 and 2346 words. Whereas an increase in ballot complexity from the 10th to the 90th percentile is associated with an increase from 2820 and 11635 words. The estimates in this section are based on the Heckman Model (4) in Table 3.

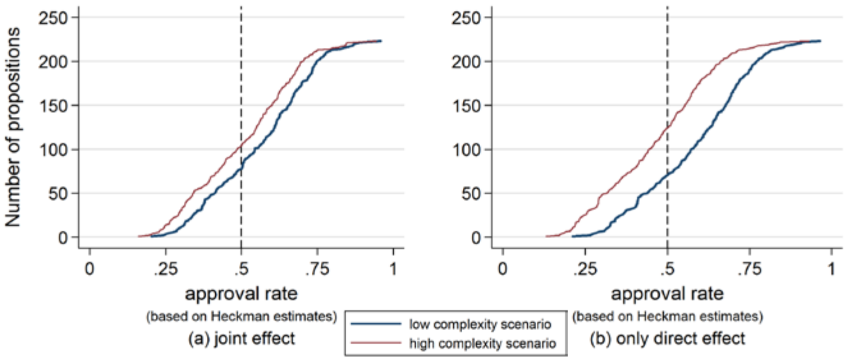


FIGURE 5. SIMULATION ON THE EFFECT OF COMPLEXITY ON APPROVAL

Notes: This figure illustrates the effect of an increase in the complexity from the 10th percentile (low complexity scenario) towards the 90th percentile (high complexity scenario). The figure is based on estimates of the approval rate for 223 propositions from 74 ballots.

In subfigure (a), the vertical distance between the red and the blue line at the 50% threshold (dashed vertical line) is 27. Hence, 27 propositions would have been rejected in the high-complexity but not in the low-complexity scenario. Subfigure (b) presents the difference in the approval rate if only the direct effect of complexity is taken into account, i.e. ignoring the offsetting participation effect. Without the participation effect, the vertical distance is 54.[35]

VI. Extensions

In this section, we first introduce a survey-based subjective complexity measure and use it to validate our previous results using the objective complexity measure (section A). In section B, we investigate whether the complexity-induced status quo bias in the voting decision is driven by the mechanisms proposed in the theoretical literature. If this were the case, we should observe that citizens with a lower education level are more likely to reject a proposition.

A. *Subjective proposition complexity*

We re-run our main estimations using an alternative complexity measure which is based on survey-based subjective perceptions of the complexity of individual propositions as

[35] Based on further analysis (not shown here), the naïve probit model strongly exaggerates the effect of complexity on the number of tilted referendum outcomes: it predicts 44 additional rejected proposals. The difference between the predicted joint effect (Heckman model) and the probit model prediction is again large (27 vs. 44 additional rejected propositions) and also statistically highly significant ($n = 223$, $t = 4.14$ and $p < 0.001$). The test is based on a paired t-test. To correct the test procedure for clustering at the ballot level, we used a block bootstrap t-test relying on 999 replications.

stated by Swiss citizens.[36] In the post-referendum VOX surveys, citizens were asked whether it was difficult for them to form an opinion about the proposition (survey question: "*Did you find it rather easy or rather difficult given the provided information to imagine the impact of a yes- or no-vote on yourself with regard to this proposition?*"). The binary variable *Difficult to Form an Opinion* is a straightforward indicator for the subjective complexity of the proposition.[37]

TABLE 6—CLOSENESS OF THE REFERENDUM OUTCOME AND PARTICIPATION (SUBJECTIVE COMPLEXITY MEASURE)

Probit coefficients reported	Dep. Var.: Participation			
	(1)	(2)	(3)	(4)
Difficult to Form an Opinion (ballot)	-0.362***	-0.345***	-0.334***	-0.325***
	(0.018)	(0.017)	(0.016)	(0.016)
Exclusion Restriction				
Closeness	-0.187	-0.121	-0.164	-0.026
	(0.158)	(0.176)	(0.163)	(0.176)
Closeness squared	-1.944***	-1.819***	-1.718***	-2.001***
	(0.629)	(0.659)	(0.628)	(0.582)
Canton dummies	No	Yes	Yes	Yes
Referendum type dummies	No	Yes	Yes	Yes
Year dummies	No	No	Yes	Yes
Policy area dummies	No	No	No	Yes
Pseudo R^2	0.138	0.143	0.161	0.168
Observations	181747	181747	181747	181747
p-value for joint significance of linear and quadratic terms in:				
Closeness	0.005	0.018	0.020	0.003

Notes: The table illustrates the significant hump-shaped effect of the closeness and the probability to turn out in the elections. Regression coefficients based on probit regression are reported in all specifications. All equations are estimated including individual controls as reported in Table 1. The variable closeness is based on data obtained from the official Swiss election data (University of Bern, Institute of Political Science 2013). *Difficult to Form an Opinion* represents the respondents' perceived difficulty to form an opinion about a particular proposition (voting equation) or in general with respect to all propositions on a specific ballot (participation equation). The table also reports the p-value for the joint significance of the variable closeness and its squared term. Heteroskedasticity-robust standard errors clustered at the ballot level are reported in parentheses. *** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level.

TABLE 7—HECKMAN SELECTION MODELS (SUBJECTIVE COMPLEXITY MEASURE)

Heckman coefficients reported	Heckman (1)		Heckman (2)		Heckman (3)		Heckman (4)	
	(1-1)	(1-2)	(2-1)	(2-2)	(3-1)	(3-2)	(4-1)	(4-2)
Dep. Variable:	Participation	Yes-Vote	Participation	Yes-Vote	Participation	Yes-Vote	Participation	Yes-Vote
Difficult to Form an Opinion (proposition)		-0.188***		-0.211***		-0.194***		-0.176***
		(0.020)		(0.030)		(0.024)		(0.023)
Difficult to Form an Opinion (ballot)	-0.327***		-0.376***		-0.365***		-0.355***	
	(0.033)		(0.021)		(0.019)		(0.018)	
Exclusion Restriction								
Closeness	-1.062***		-0.501		-0.447*		-0.189	
	(0.165)		(0.322)		(0.234)		(0.255)	

[36] We do not use the subjective complexity measure in section V because it has several shortcomings compared to our objective complexity measure. First, it is potentially endogenous. For instance, non-participants may ex post justify their absenteeism with the excuse that it was difficult to decide. A second shortcoming arises due to the correlation of subjective complexity with citizens' characteristics (see section VI.B for more details).

[37] In the participation equation, we use the more general survey question: *"In general, did you find it rather easy or rather difficult given the provided information to imagine the impact of a yes- or no-vote on yourself?"*

Closeness squared	-2.153***	-1.961***	-1.683**	-2.151***
	(0.603)	(0.747)	(0.809)	(0.749)
Unobserved Factors				
rho	0.879	0.388	0.299	0.233
Wald test (*p-value*)	0.000	0.048	0.005	0.011
Canton dummies	No	Yes	Yes	Yes
Referendum type dummies	No	Yes	Yes	Yes
Year dummies	No	No	Yes	Yes
Policy area dummies	No	No	No	Yes
Observations	166787	166787	166787	166787

Notes: The table provides the estimated coefficients of the Heckman selection model and establishes the negative and significant effect of complexity on voter's probability to participate and vote in favor of a proposition. The table also reports the correlation ρ between the error terms of both equations, as well as the corresponding p-values. *Difficult to Form an Opinion* represents the respondents' perceived difficulty to form an opinion about a particular proposition (voting equation) or in general with respect to all propositions on a specific ballot (participation equation). All equations are estimated with individual controls as reported in Table 1, fixed effects for the ballot year, referendum type and the canton in which the eligible voter lives. Heteroskedasticity-robust standard errors clustered at the ballot level are reported in parentheses. *** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level.

All estimation results are qualitatively in line with our previous results reported in section V.B. Table 8 provides the corresponding marginal effects related to the Heckman estimations and the comparisons to the results from simple probit estimations.

TABLE 8—PARTIAL EFFECTS OF HECKMAN MODELS (SUBJECTIVE COMPLEXITY MEASURE)

APE	Probit	Probit	Heckman		
	π=1	v=1|π=1	π=1	v=1|π=1	v=1
Difficulty to Form an Opinion (proposition)		-0.056***		-0.065***	-0.063***
		(0.008)		(0.009)	(0.008)
Difficulty to Form an Opinion (ballot)	-0.116***		-0.115***	0.017***	
	(0.006)		(0.006)	(0.006)	

Notes: The table compares the average partial effects of a discrete change in each complexity measure on participation and voting behavior based on the single equation models and the Heckman selection model (column (4-1) and (4-2) in Table 7). The Heckman model allows for an indirect effect of the variable subjective complexity measure (ballot) on the vote outcome via altering the participation decision. This indirect effect increases the probability of voting in favor of a proposition by 1.7 percentage points. Even though quantitatively important the indirect effect is outweighed by the negative direct effect of subjective complexity on the probability of voting in favor of a proposition of -6.5 percentage points. *** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level.

Voters who find it rather difficult to form an opinion are 11.5 ppts less likely to turn out and about 4.7 ppts (6.5 ppts - 1.7 ppts) more likely to reject a proposition. Similar to the results related to Table 5 in section V.B, relying on simple probit models would lead to an overestimation of the complexity effect.

The results estimated with the subjective complexity measure are qualitatively similar to the objective complexity estimates and provide additional evidence that our booklet-based objective complexity measure is indeed a valid measure of the underlying complexity of a proposition and that complexity has a considerable effect on individual voting behavior in referendums.[38]

[38] Note that since the subjective measure also varies within propositions (i.e. at the respondent level) one could also include proposition fixed effects instead of year, referendum type and topic fixed effects in the participation equation. We have conducted additional estimations along these lines within a two-equation probit framework (results available upon request) and the results are virtually identical.

B. *Complexity and education*

The availability of an objective complexity measure is particularly valuable when interacting proposition complexity with individual voter characteristics such as education. Let $c_s(c_o,\ \mathbf{i_{obs}}, \mathbf{i_{uno}})$ denote the survey-based subjective complexity of a proposition which depends on the objective complexity of the proposition c_o, a vector of observable individual characteristics $\mathbf{i_{obs}}$ and a vector of unobservable individual characteristics (such as intelligence or cognitive skills) denoted by $\mathbf{i_{uno}}$. Highly educated people may be better able to deal with complex issues leading to differences in voters' reaction to complex propositions across education levels.[39] Voter's education is, however, also likely to be correlated with unobservable characteristics $\mathbf{i_{obs}}$. The corresponding interaction term between subjective complexity and education is $c_s(c_o, edu,\ \mathbf{i_{obs}}, \mathbf{i_{uno}}) \times edu$. It is, however, not clear how to interpret the estimate for this interaction term because as education changes, the perceived subjective complexity c_s changes as well.[40] We circumvent the problems related to a subjective complexity measure and use our objective text-based measure of complexity which is uncorrelated with individual voter characteristics. Thus, we estimate $c_o \times edu$ which has a clear interpretation.

Based on our considerations in Section II, we expect that with increasing complexity, less educated voters are more likely to use the status-quo heuristic. To test this, we re-estimate the Heckman selection model that interacts proposition complexity and a university degree dummy which equals 1 for citizens with a university degree and 0 otherwise:[41]

$$Participate^* = \alpha\ Complexity + \lambda\ University + \gamma\ Complexity \times University + \boldsymbol{\beta}'\mathbf{x}\ + \boldsymbol{\delta}'\mathbf{z} + u,$$
$$Participate = 1 \text{ if } Participate^* > 0, Participate = 0 \text{ otherwise},$$

(5)
$$Yes\text{-}vote^* = \alpha\ Complexity + \lambda\ University + \gamma\ Complexity \times University + \boldsymbol{\beta}'\mathbf{x} + \epsilon,$$
$$Yesvote = 1 \text{ if } Yesvote^* > 0, Yesvote = 0 \text{ otherwise}.$$

Table 9 presents the estimation results. The upper part of the table reports regression coefficients; the lower part reports the marginal effect for the variables of main interest. Model (1) in Table 9 provides the estimation results that we obtain when we do not control for

[39] We indeed find that subjective complexity is positively related to objective complexity and negatively related to education. Table A.8 in the online appendix provides the empirical results.

[40] The only way how the level of subjective complexity can stay constant w.r.t. to a change of complexity is if we assume a simultaneous change in unobservable characteristics offsetting the effect of education on the level of subjective complexity.

[41] Distinguishing citizens' education levels only w.r.t to university and non-university degree keeps the analysis tractable and ensures that we can interpret our education unambiguously w.r.t. high and low education.

education.[42] A university degree dummy is introduced in Model (2) and (3). Citizens with a university degree have an ~11 ppts higher probability to participate in a referendum and a ~6 ppts higher probability to vote in favor of a proposition. The coefficient for the interaction term between university education and complexity is positive but insignificant. However, the size, the sign and the statistical significance cannot be interpreted for interaction effects in nonlinear models (Ai and Norton 2003; Greene 2010; Berry, DeMeritt, and Esarey 2010).[43] To quantify the size and statistical significance of the interaction effect we follow the suggestion by Greene (2010) and analyze the predicted probabilities of participation and yes-voting for citizens with and without a university degree. With regard to the participation decision, we find no meaningful heterogeneity (figure available upon request).

TABLE 9—STATUS QUO BIAS AND THE INTERACTION BETWEEN COMPLEXITY AND EDUCATION

	Heckman (1)		Heckman (2)		Heckman (3)	
	(1-1)	(1-2)	(2-1)	(2-2)	(3-1)	(3-2)
Dep. Variable:	Participation	Yes-Vote	Participation	Yes-Vote	Participation	Yes-Vote
Complexity (proposition)		-0.344*** (0.083)		-0.343*** (0.082)		-0.353*** (0.082)
Complexity (ballot)	-0.527*** (0.121)		-0.517*** (0.120)		-0.516*** (0.121)	
University degree			0.341*** (0.025)	0.217*** (0.035)	0.390 (0.399)	-0.521 (0.537)
Complexity (proposition) x Uni						0.100 (0.073)
Complexity (ballot) x Uni					-0.006 (0.047)	
Average Marginal Effects	Pr(π=1)	Pr(v=1|π=1)	Pr(π=1)	Pr(v=1|π=1)	Pr(π=1)	Pr(v=1|π=1)
Complexity (proposition)		-0.128*** (0.033)		-0.128*** (0.032)		-0.128*** (0.032)
Complexity (ballot)	-0.168*** (0.038)	0.039*** (0.013)	-0.164*** (0.037)	0.041*** (0.013)	-0.164*** (0.037)	0.041*** (0.013)
University degree			0.106*** (0.008)	0.057*** (0.011)	0.106*** (0.008)	0.053*** (0.011)
Observations	191669		191669		191669	

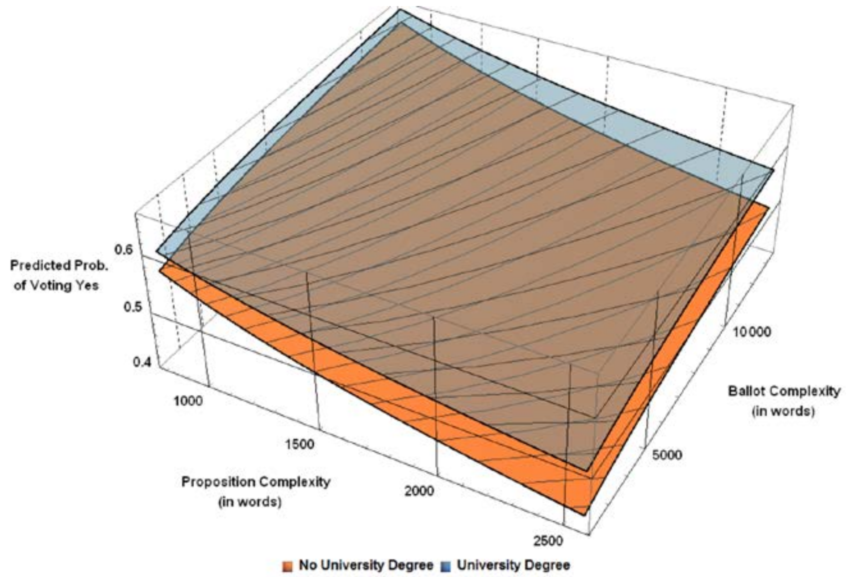
Notes: The table reports the estimates of the interaction effect between the objective complexity measure and education (university degree vs. no university degree) and therefore indicates heterogeneity in the response to complexity for voters with different education levels. Heckman coefficients are reported in the upper half of the table. Average marginal effects are reported in the lower half of the table. The average marginal effect associated with the interaction term is illustrated in Figure 6. *Complexity* represents the log of the number of words used in the information text in the official booklets per proposition (voting equation) or aggregated at the ballot level (participation equation). All regressions are estimated with fixed effects for year, canton, policy area, and referendum type and controls for individual characteristics. Heteroskedasticity-robust standard errors clustered at the ballot level are reported in parentheses. *** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level.

Figure 6 presents the results for the voting decision. The slopes of the two surfaces in subfigure (a) of Figure 6 illustrate the marginal effect of complexity on the likelihood of accepting a proposition conditional on voters' education level. Subfigure (a) shows that there

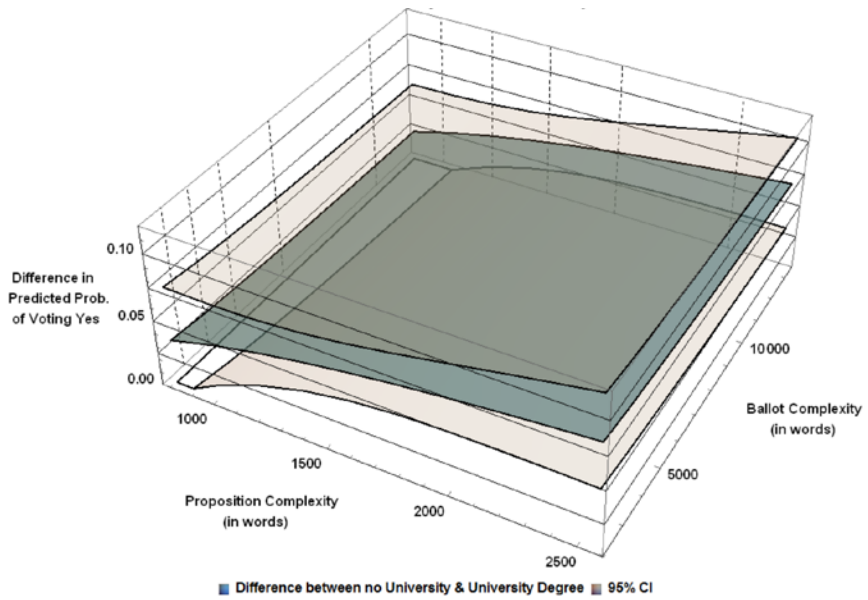
[42] Thus, Model (1) in Table 6 only differs from Model (4) in Table 3 in the sense that we do not control for education.

[43] Hence, using a t-test to assess statistical significance for the coefficient of the interaction term is also invalid.

is status quo bias regardless of the education level of voters. Low-educated voters reject propositions more often than high-educated voters even when propositions have a low level of complexity. This difference between low-educated and high-educated voters increases with proposition complexity which is illustrated more clearly in subfigure (b). The difference in the expected probability of voting in favor of a proposition increases with proposition complexity from below 3% to about 8% and is almost always significant.



(a)



(b)

FIGURE 6. PREDICTED PROBABILTIES OF „YES-VOTE“ - VOTERS WITH AND WITHOUT UNIVERSITY DEGREE

Notes: Figure 6 is based on the estimates of Model (4) in Table 9. Dotted vertical lines correspond to the 10th, 25th, 50th, 75th, and 90th percentiles of the respective complexity measure. For better readability, the complexity axes are restricted to values between the 5th and 95th percentile. Figure (a) illustrates the heterogeneous effect of complexity on voters' probability to vote in favor or against a proposition for two groups of voters. Voters without a university degree change their voting behavior stronger towards the status-quo as complexity rises. The figure plots the average predicted probability against the complexity of the proposition (direct effect) and the complexity of the ballot (indirect effect). Figure (b) illustrates the "difference in difference" w.r.t. the response to complexity of citizens with and without a university degree.

VII. Robustness tests

In this section, we discuss the results for five robustness tests. We only report and interpret the main findings here. Tables and figures are available in the online appendix.[44]

A. *Does it matter whether voters read the information booklet?*

We argue that the length of the text describing a proposition in the official information booklet serves as proxy for the complexity of the proposition, independently of whether the voter has actually seen or bothered to read the booklet.[45] An alternative mechanism which may explain the link between the length of the information text and the complexity of the proposition works as follows: If voters read a complex description of a proposition in the information booklet, they might be overwhelmed by the length of the information text itself. This would open up the possibility that a voter is overwhelmed not because the underlying proposition is complex, but rather because the description of the proposition is complex.

If this alternative mechanism drives our main estimation results, including a dummy that indicates whether a citizen has used the relevant information booklet should significantly affect our estimates for the effect of complexity on individual voting behavior. A second empirical implication of this potential mechanism is that the length of the information text should alter individual voting behavior only for those voters who actually use the information booklet. We test this mechanism by estimating the following model:

(6) $$Participate^{*} = \alpha\ Complexity + \lambda\ Booklet + \gamma\ Complexity \times Booklet + \boldsymbol{\beta}'\mathbf{x} + \boldsymbol{\delta}'\mathbf{z} + u,$$
$$Participate = 1 \text{ if } Participate^{*} > 0, Participate = 0 \text{ otherwise,}$$

(7) $$Yes\text{-}vote^{*} = \alpha\ Complexity + \lambda\ Booklet + \gamma\ Complexity \times Booklet + \boldsymbol{\beta}'\mathbf{x} + \epsilon,$$
$$Yesvote = 1 \text{ if } Yesvote^{*} > 0, Yesvote = 0 \text{ otherwise,}$$

where *Booklet* is a dummy variable that indicates whether a voter has used the information booklet. If the alternative channel is relevant, the effect of our complexity measure on the status quo bias should diminish or at least decrease substantially.

Table A.4 in the online appendix presents the estimation results. The lower part of the table reports average marginal effects. The inclusion of the *booklet* dummy does not affect the joint effect of proposition and ballot complexity on the likelihood to vote against a

[44] In an additional robustness test (not reported here), we rerun the estimations for mandatory and optional referendums only (i.e. we exclude initiatives and counterinitiatives). The results are qualitatively similar. Separate estimations for initiatives are not feasible given the small number of observations.

[45] The booklets provide a reputable and widely used information source for a majority of voters (Rohner 2012).

proposition (Model (2)). Model (3) includes the interaction term.[46] These effects are independent of proposition complexity and are therefore not directly linked to the effect of proposition complexity on the vote outcome. As mentioned before, an insignificant point estimate for the interaction term does not indicate a nonlinear interaction effect between proposition complexity and booklet use (Greene 2010). Therefore, we plot the predicted probabilities of participation (Figure A.2 in the online appendix) and supporting a proposition (Figure A.3 in the online appendix) for booklet readers and nonreaders. The results show that regardless of whether voters have read or not read the booklet, proposition complexity has a similar effect on voting behavior. We conclude that the alternative channel is not empirically relevant and that the underlying complexity of a proposition matters for individual voting behavior.

B. *Proposition complexity and proposition importance*

More important propositions may be associated with longer booklet texts. Therefore, our complexity measure may be confounded. If so, the question arises whether importance and not complexity is the mechanism that drives the effect of our complexity measure on voting behavior. Voters may be more likely to participate in a referendum if they perceive the proposition to be important. In addition, the importance of a proposition likely mitigates the status quo bias because voters should be more likely to invest resources in understanding complex propositions when they perceive them as important. If our text-based complexity measure is indeed confounded with the importance of the proposition, our estimates should be biased towards zero, i.e. the true effect of complexity on the probability of abstaining from voting and rejecting a proposition is larger in absolute terms. To check for this possibility, we estimate the following models:

(8) $$Participate^{*} = \alpha \, Complexity + \lambda \, Importance + \boldsymbol{\beta}'\mathbf{x} + \boldsymbol{\delta}'\mathbf{z} + u,$$
$$Participate = 1 \text{ if } Participate^{*} > 0, Participate = 0 \text{ otherwise},$$

(9) $$Yes\text{-}vote^{*} = \alpha \, Complexity + \lambda \, Importance + \boldsymbol{\beta}'\mathbf{x} + \epsilon,$$
$$Yesvote = 1 \text{ if } Yesvote^{*} > 0, Yesvote = 0 \text{ otherwise},$$

[46] These results also provide interesting evidence of an effect that is not important for the mechanism that is tested here but which may be of interest to readers. Voters using the information booklet are 2 ppts more likely to vote in favor of a proposition and almost 18 ppts more likely to participate in a referendum (see Models (2) and (3)) in Table A.4. This shows that when voters make an effort of collecting more information they are more likely to turn out and less likely to reject a proposition. This is in line with previous theoretical considerations on the role of information for voting behavior (Matsusaka 1995; Feddersen and Pesendorfer 1996), even though of course the decision to read the booklet is endogenous.

where *Importance* is a categorical dummy variable (scaled from 0 (unimportant) to 10 (highly important) measuring two types of importance that a survey respondent attaches to a proposition. In particular, the VOX survey asks the following questions: "*How important are the consequences of the proposition for you personally?*" and "*How important are the consequences of the proposition for our country?*"

Since the survey questions about the perceived proposition importance were only asked in a subset of the referendums in our sample, we re-estimate our baseline model in column (1) in Table A.5 in the online appendix to obtain a benchmark with this smaller sample. The estimates are not substantially affected by the drop in sample size. When we control for proposition importance, the negative effect of (objective) complexity on voting behavior changes only slightly (see columns (2) and (3): from -0.14 to -0.157 and from -0.137 to -0.161). This suggests that proposition complexity, and not proposition importance, drives the negative effect. The average marginal effect of proposition importance on the probability of voting in favor of a proposition is positive and highly significant. We conclude that while proposition importance makes it more likely that voters turn out and support a proposition, our complexity measure does not seem to be confounded by proposition importance.

C. *Endogeneity of proposition complexity: Strategic manipulation by the government?*

Another concern is that the government may manipulate the information text of a proposition to influence citizens' voting behavior. Note however that the scope for manipulation by changing the booklet text is limited due to legal restrictions on the booklets (see section IV.B) as well as the absence of systematic differences in the effect of complexity on voting behavior between booklet users and non-users (see our results in section VII.B). Nevertheless, in this section we investigate whether the complexity of propositions (as measured by the length of the information text) is systematically smaller (larger) when the Swiss government supports (is against) a proposition. We estimate the following model:

$$Complexity = \alpha\ GovernmentAgainst + \boldsymbol{\beta}'\mathbf{x} + u\ , \tag{10}$$

where *GovernmentAgainst* is a dummy variable that is 1 when the National Council advises voters to reject a proposition and 0 when the National Council advises voters to support a proposition. The data for this variable is taken from official election data provide by the University of Bern, Institute of Political Science (2013). If indeed the government attempts to manipulate voters, we expect that the estimate for α is positive and significant.

The estimation results are reported in Table A.6 in the online appendix. We find that there is no significantly positive correlation between the government's voting recommendation and our complexity measure. We conclude that our complexity measure is not confounded by attempts of the government to influence referendum outcomes in its favor.[47]

D. *Exclusion restriction based on pre-poll data*

Ideally we would use data on the ex-ante closeness of referendum outcomes for the exclusion restriction. Pre-poll data based on official surveys conducted by gfs.bern is officially available as of 2004. To increase data coverage, we have additionally searched through archives of major Swiss newspapers and found a few additional pre-polls prior to 2004.[48] In total, we have been able to collect pre-poll data for 47 out of the 223 referendums in our sample. This reduced sample would be too small for reliable statistical inference in a Heckman selection model.[49] Therefore, we have used this data to impute a measure of ex-ante closeness for the entire sample.[50] The estimation results are collected in Tables A.2 and A.3 in the online appendix. Table A.2 summarizes the results for the prediction. Table A.3 reports the Heckman estimation results based on the ex-ante closeness measure.

The prediction in Table A.2 fits the data quite well: in the most complete specification (Model (4)) the R^2 for the imputation is 0.90, the partial correlation between the ex-ante and ex post closeness amounts to 0.82 and is highly significant (t-statistic of 5.4). The Heckman estimations in Table A.3 provide results that are comparable to those in our baseline estimations. In particular, we obtain a hump-shaped relationship between closeness and the likelihood of participation (significant at 10% level) and a highly significant selection bias.

---

[47] Note that our results do not prove that the government has no power at all to influence voting decision in a referendum. However, our results show that such manipulation does not appear to occur via the complexity of propositions. See Selb (2008) for details on the limits of Swiss national government to influence the composition and content of ballots.

[48] The only other study that we are aware of using this pre-poll data is Bursztyn et al. (2017) which shows that elections that are forecast to be close are indeed associated with a higher turnout.

[49] In fact, with this small sample we still obtain qualitatively similar results, i.e. we obtain an inverse U-shape for the relationship between the ex-ante closeness and the likelihood of participation. However, the standard errors are fairly large leading to a t-statistic of about 1.6 for the squared term.

[50] We calculated predicted values for ex-ante closeness as follows. First, we regressed the measure for ex-post closeness on the measure for ex-ante closeness and various controls for the subset of 47 referendums where the ex-ante measure is available. Then we used the estimated coefficients to predict an ex-ante closeness measure for the entire sample of 223 referendums. Finally, we re-ran the baseline estimations in Table 3 using this new closeness measure. Since 77% of the predicted values are out-of-sample predictions, these additional results should be viewed as suggestive evidence to support our baseline estimates.

E. *Potential survey bias in post-referendum surveys*

A general concern with post-election surveys is survey bias, i.e. voters' responses may not be truthful. Comparing the aggregate results of the VOX survey and official election data, Funk (2016) provides evidence for a significant difference in the share of yes-votes in about half of the referendums. She also indicates in her paper which propositions are affected by survey bias.[51] This allows us to test whether our estimation results on the effect of complexity on individual voting behavior may be contaminated by survey bias. Due to a sense of civic duty, voters may feel pressured to state that they participated in a referendum when they actually did not. It is, however, ex ante unclear how this may relate to the effect of complexity on turnout or especially on the actual voting decision. The main purpose of this exercise is to obtain somewhat "cleaner" estimates in our baseline models by excluding those propositions where Funk (2016) provides evidence for survey bias.

We re-estimate our baseline Heckman model for different subsamples, in which we exclude propositions with the highest survey bias as identified in Funk (2016). Table A.7 in the online appendix summarizes the estimation results. In columns (1) to (4), we report estimates for different subsamples excluding 5%, 10%, 25%, and 50% of the propositions with the highest survey bias as reported in Funk (2016). Even if we exclude 50% of the propositions with the highest survey bias our results remain fairly unaffected. We conclude that there is no systematic influence of survey bias on the effect of complexity on individual voting behavior.

VIII. Conclusion

One concern that is often voiced by scholars about direct democracy is that citizens may not be sufficiently competent to decide about complex policy issues. This paper is the first to study how the complexity of propositions affects individual voting behavior in a direct democracy using a Heckman selection approach. Our dataset combines a novel complexity measure based on information provided in official booklets with individual post-referendum

[51] Note that our objective is to explain voting behavior at the individual level rather than predicting exact aggregate referendum outcomes. Whenever we rely on aggregate data in our estimations we use official election data. In the estimations in Table 2, we rely on the closeness of the referendum outcome to identify the Heckman selection model, where we took the data from the official election data from the University of Bern, Institute of Political Science (2013). Funk (2016) also shows that the survey bias varies across specific proposition topics. We include topic fixed effects meaning we rely on differences in the probability of voting yes within each topic category. Our point estimates usually increase and become more precise with topic fixed effects, possibly due to control for biases mentioned in Funk (2016).

survey data for 223 referendums at the federal level in Switzerland over the 1981-2010 period.

We find that the more complex a proposition is, the less likely are citizens to participate and, if they participate, to support a proposition. These findings are consistent with the idea of a status quo bias when issues are too complex. More educated voters respond less to increasing complexity and exhibit a lower status quo bias than less educated voters.

Two competing mechanisms determine the impact of complexity on the status quo bias in the vote outcome. A *direct effect* – confronted with higher complexity, voters tend to vote in favor of the status quo, and an *indirect participation effect* – increasing complexity reduces the percentage of biased citizens (through vote abstentions) among the voters and therefore mitigates the status quo bias. We contribute to the literature by estimating the magnitude of both effects. We find that neglecting the indirect participation effect would overestimate the effect of complexity on the referendum outcome by almost 50%. The sheer magnitude of the opposing indirect effect calls attention to the importance of treating voting behavior as an outcome of two sequential choices.

The normative assessment of the decline in turnout triggered by proposition complexity depends on the trade-off between a *representative vote outcome* (Lijphart 1997) and an *informationally superior vote outcome* (Feddersen and Pesendorfer 1996). We provide evidence for the latter mechanism suggesting that policy measures like the introduction of mandatory voting may be counterproductive. Finally, our results suggest that improving the general level of education seems to be an appropriate measure to reduce the bias in the vote outcome.

References

Ai, Chunrong and Edward C. Norton. 2003. "Interaction terms in logit and probit models." *Economics Letters*, 80(1): 123–29.

Berry, William D., Jacqueline H. R. DeMeritt, and Justin Esarey. 2010. "Testing for Interaction in Binary Logit and Probit Models: Is a Product Term Essential?" *American Journal of Political Science*, 54(1): 248–66.

Bowler, Shaun and Todd Donovan. 1998. *Demanding choices. Opinion, voting, and direct democracy*. Ann Arbor: University of Michigan Press.

Bowler, Shaun, Todd Donovan, and Trudi Happ. 1992. "Ballot Propositions and Information Costs: Direct Democracy and the Fatigued Voter." *The Western Political Quarterly*, 45(2): 559–68.

Brunner, Eric, Stephen L. Ross, and Ebonya Washington. 2011. "Economics and Policy Preferences: Causal Evidence of the Impact of Economic Conditions on Support for Redistribution and Other Ballot Proposals." *Review of Economics and Statistics*, 93(3): 888–906.

Bundesgericht. 2008. *Bundesgerichtsurteil 1C_412/2007 E.6.1*.

Bursztyn, Leonardo, Davide Cantoni, Patricia Funk, and Noam Yuchtman. 2017. "Polls, the Press, and Political Participation. The Effects of Anticipated Election Closeness on Voter Turnout." *National Bureau of Economic Research Working Paper Series*, No. 23490.

Cox, Gary W. and Michael C. Munger. 1989. "Closeness, Expenditures, and Turnout in the 1982 U.S. House Elections." *American Political Science Review*, 83(01): 217–31.

Cronin, Thomas E. 1999. *Direct democracy. The politics of initiative, referendum, and recall*. Cambridge, Mass: Harvard University Press.

Denver, David T. and H. T. G. Hands. 1974. "Marginality and Turnout in British General Elections." *British Journal of Political Science*, 4(01): 17–35.

Downs, Anthony. 1957. *An economic theory of democracy*. New York: Harper and Row.

Eichenberger, Reiner and Angel Serna. 1996. "Random errors, dirty information, and politics." *Public Choice*, 86(1): 137–56.

Feddersen, Timothy J. and Wolfgang Pesendorfer. 1996. "The Swing Voter's Curse." *The American Economic Review*, 86(3): 408–24.

FORS - Swiss foundation for research in social sciences. 2012. "Voxit data set." University of Lausanne.

Freedman, David A. and Jasjeet S. Sekhon. 2010. "Endogeneity in Probit Response Models." *Political Analysis*, 18(2): 138–50.

Frey, Bruno S. 1994. "Direct Democracy: Politico-Economic Lessons from Swiss Experience." *American Economic Review*, 84(2): 338–42.

Funk, Patricia. 2016. "How Accurate Are Surveyed Preferences for Public Policies? Evidence from a Unique Institutional Setup." *Review of Economics and Statistics*, 98(3): 442–54.

Funk, Patricia and Christina Gathmann. 2013. “Voter preferences, direct democracy and government spending.” *European Journal of Political Economy*, 32(0): 300–19.

Funk, Patricia and Christina Gathmann. 2015. “Gender Gaps in Policy Preferences: Evidence from Direct Democracy in Switzerland.” *Economic Policy*, 30(81): 141–81.

Gerber, Elisabeth R. and Arthur Lupia. 1999. “Voter Competence in Direct Legislation Elections.” In *Citizen competence and democratic institutions*, ed. Stephen L. Elkin and Karol E. Sołtan. University Park, Pa.: Pennsylvania State University Press.

Gigerenzer, Gerd and Wolfgang Gaissmaier. 2010. “Heuristic Decision Making.” *Annual Review of Psychology*, 62(1): 451–82.

Greene, William. 2010. “Testing hypotheses about interaction terms in nonlinear models.” *Economics Letters*, 107(2): 291–96.

Heckman, James J. 1978. “Dummy Endogenous Variables in a Simultaneous Equation System.” *Econometrica*, 46(4): 931–59.

Heckman, James J. 1979. “Sample Selection Bias as a Specification Error.” *Econometrica*, 47(1): 153–61.

Hessami, Zohal. 2016. “How do voters react to complex choices in a direct democracy? Evidence from Switzerland.” *Kyklos*, 62(2): 263–93.

Hofer, Katharina E., Christian Marti, and Monika Bütler. 2017. “Ready to reform: How popular initiatives can be successful.” *European Journal of Political Economy*, 48: 16–39.

Huber, John D. and Charles R. Shipan. 2002. *Cambridge studies in comparative politics, Deliberate discretion? The institutional foundations of bureaucratic autonomy / John D. Huber, Charles R. Shipan*. Cambridge: Cambridge University Press.

Key, Valdimer O. and Alexander Heard. 1984. *Southern politics in state and nation*. Knoxville: University of Tennessee Press.

Kriesi, Hanspeter. 2008. *Direct democratic choice. The Swiss experience*. 1st ed. Lanham, Md: Lexington, Books.

Krishnakumar, Jaya and Tobias Müller. 2012. “The political economy of immigration in a direct democracy: The case of Switzerland.” *European Economic Review*, 56(2): 174–89.

Lapalombara, Joseph G. and Charles B. Hagan. 1951. “Direct Legislation: An Appraisal and a Suggestion.” *American Political Science Review*, 45(02): 400–21.

Lijphart, Arend. 1997. “Unequal Participation: Democracy's Unresolved Dilemma Presidential Address, American Political Science Association, 1996.” *American Political Science Review*, 91(01): 1–14.

Lupia, Arthur. 1994. “Shortcuts versus Encyclopedias: Information and Voting Behavior in California Insurance Reform Elections.” *American Political Science Review*, 88(1): 63–76.

Magleby, David B. 1984. *Direct legislation. Voting on ballot propositions in the United States*. Baltimore: Johns Hopkins University Press.

Matsusaka, John G. 1995. “Explaining voter turnout patterns: An information theory.” *Public Choice*, 84(1): 91–117.

Matsusaka, John G. and Filip Palda. 1993. “The Downsian voter meets the ecological fallacy.” *Public Choice*, 77(4): 855–78.

McCloskey, Deirdre N. and Stephen T. Ziliak. 1996. “The Standard Error of Regressions.” *Journal of Economic Literature*, 34(1): 97–114.

Meier, Armando N., Lukas D. Schmid, and Alois Stutzer. 2016. “Rain, Emotions and Voting for the Status Quo.” *IZA Discussion Papers*(10350).

Nicholson, Stephen P. 2003. “The Political Environment and Ballot Proposition Awareness.” *American Journal of Political Science*, 47(3): 403.

Reilly, Shauna and Sean Richey. 2011. “Ballot Question Readability and Roll-Off: The Impact of Language Complexity.” *Political Research Quarterly*, 64(1): 59–67.

Riker, William H. and Peter C. Ordeshook. 1968. “A Theory of the Calculus of Voting.” *The American Political Science Review*, 62(1): 25–42.

Rohner, Gabriela. 2012. *Schriften zur Demokratieforschung*. Vol. 4, *Die Wirksamkeit von Volksinitiativen im Bund, 1848-2010*. Zürich [etc.]: Schulthess.

Samuelson, William and Richard Zeckhauser. 1988. “Status Quo Bias in Decision Making.” *Journal of Risk & Uncertainty*, 1(1): 7–59.

Schweizer Bundesrat. 1978. Vol. 130. *Bundesratsbeschluss über das Ergebnis der Volksabstimmung vom 4. Dezember 1977 (Politische Rechte; Ausgleich des Bundeshaushalts). BBI*(I): 324.

Selb, Peter. 2008. “Supersized Votes: Ballot Length, Uncertainty, and Choice in Direct Legislation Elections.” *Public Choice*, 135(3-4): 319–36.

Stadelmann, David and Benno Torgler. 2013. “Bounded Rationality and Voting Decisions over 160 Years: Voter Behavior and Increasing Complexity in Decision-Making.” *PLoS ONE*, 8(12): e84078.

University of Bern, Institute of Political Science. 2013. “Datensatz der eidgenössischen Volksabstimmungen ab 1848.” Année Politique Suisse. http://www.swissvotes.ch/uploads/codebuch.pdf.

Van de Ven, Wynand P.M.M. and Bernard M.S. van Praag. 1981. “The demand for deductibles in private health insurance: A probit model with sample selection.” *Journal of Econometrics*, 17(2): 229–52.

Wooldridge, Jeffrey M. 2004. “Statistical significance is okay, too: comment on “Size Matters”” *The Journal of Socio-Economics*, 33(5): 577–79.

Wooldridge, Jeffrey M. 2010. *Econometric analysis of cross section and panel data*. 2nd ed. Cambridge, Mass: MIT Press.

Ziliak, Stephen T. and Deirdre N. McCloskey. 2008. *Economics, cognition, and society, The cult of statistical significance. How the standard error costs us jobs, justice, and lives*. Ann Arbor: University of Michigan Press.

Online Appendix

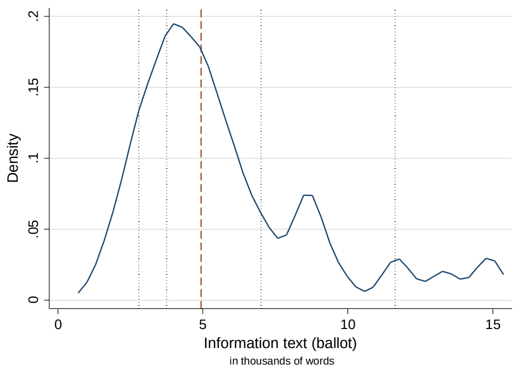


FIGURE A.1. VARIATION IN THE BOOKLET-BASED COMPLEXITY MEASURE

Notes: This figure depicts the variation in the booklet-based objective complexity measure aggregated at the ballot level. It is based on a Gaussian kernel density plot with a half-width of 500 words. The dashed red line indicates the median. The dotted grey lines correspond to the 10th , 25th , 75th , and 90th percentiles. For better readability, the complexity axis is restricted to values below the 99th percentile.

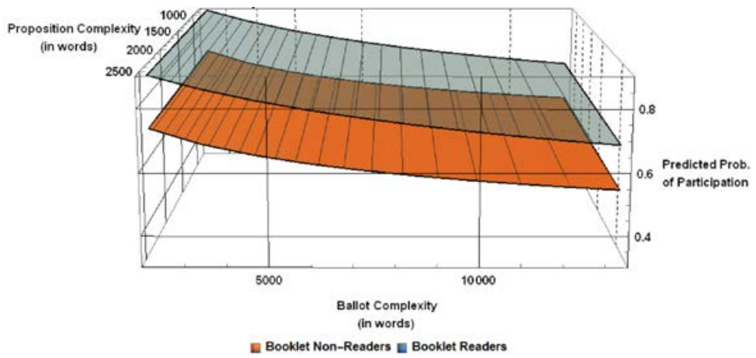


FIGURE A.2. PREDICTED PROBABILTY OF PARTICIPATION AND THE BOOKLET READING CHANNEL

Notes: This figure illustrates that booklet readers and non-readers participation behavior is fairly similar w.r.t. their reaction to complexity. However, booklet readers are more likely to participate in the election regardless of the ballot complexity. The figure plots the average predicted probability against the complexity of the proposition (direct effect) and the complexity of the ballot (indirect effect). The estimates in Figure A.2 are based on the estimates of Model (3) in Table A.4. The dotted vertical lines correspond to the 10th, 25th, 50th, 75th, and 90th percentile of the respective complexity measure. For better readability, the complexity axes are restricted to values between the 5th and 95th percentile.

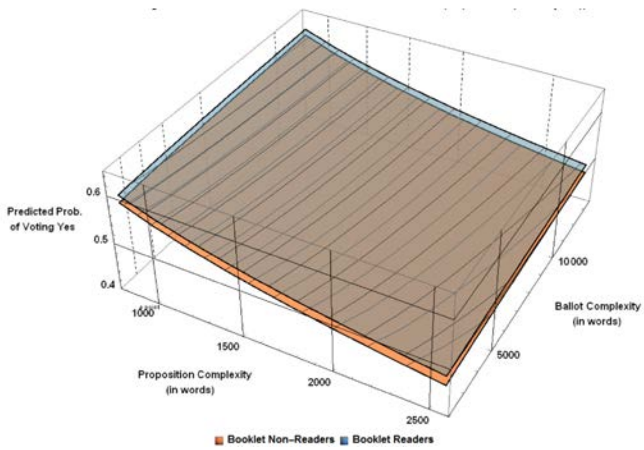


FIGURE A.3. BOOKLET READING, COMPLEXITY AND VOTING BEHAVIOR

Notes: This figure illustrates that booklet readers and non-readers voting behavior is fairly similar w.r.t. their reaction to complexity. The figure plots the average predicted probability against the complexity of the proposition (direct effect) and the complexity of the ballot (indirect effect). Figure A.3 is based on the estimates of Model (3) in Table A.4. The dotted vertical lines correspond to the 10^{th}, 25^{th}, 50^{th}, 75^{th}, and 90^{th} percentile of the respective complexity measure. For better readability, the complexity axes are restricted to values between the 5^{th} and 95^{th} percentile.

TABLE A.1—DESCRIPTIVE STATISTICS

Voters and non-voters (N=191669)				Only voters (N=106817)			
	Mean	SD	Median		Mean	SD	Median
Log Complexity (ballot)	8.58	0.52	8.5	Log Complexity (proposition)	7.33	0.44	7.3
Complexity (ballot)	6.10	3.38	5.1	Complexity (proposition)	1.73	1.55	1.5
Rural	0.35	0.48	1	Rural	0.35	0.48	1
Female	0.50	0.50	1	Female	0.47	0.50	0
Age	47.00	17.39	44	Age	49.85	16.79	49
Education	2.71	1.54	2	Education	2.93	1.61	2
Proposition Knowledge (ballot)	4.80	2.98	4	Proposition Knowledge	1.62	0.60	2
Married	0.59	0.49	1	Married	0.65	0.48	1
Protestant	0.43	0.49	0	Protestant	0.45	0.50	0
Employed	0.61	0.49	1	Employed	0.60	0.49	1

Notes: The table reports the descriptive statistics for the sample used in regressions in Table 3. The logs of the complexity measures are used in the estimations. We also report the descriptive statistics for the untransformed complexity measures (text length in thousands of words).

TABLE A.2—EX ANTE AND EX POST CLOSENESS OF THE VOTE OUTCOME

OLS coefficients reported	(1)	(2)	(3)	(4)
Dep. Variable:	Pre-poll closeness	Pre-poll closeness	Pre-poll closeness	Pre-poll closeness
Ex-post closeness	0.797***	0.718***	0.776***	0.819***
	(0.084)	(0.125)	(0.099)	(0.152)
Canton dummies	No	Yes	Yes	Yes
Referenda type dummies	No	Yes	Yes	Yes
Year dummies	No	No	Yes	Yes
Policy area dummies	No	No	No	Yes
R-squared	0.65	0.66	0.78	0.90
Observations	44454	44454	44454	44454

Notes: The table establishes a strong and significant correlation between ex post (based on pre-poll data) and ex ante (based on administrative election data) measures of closeness. All equations are estimated including individual controls as reported in Table 1. The table reports OLS regressions with heteroskedasticity-robust standard errors clustered at the referendum level in parentheses. *** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level.

TABLE A.3—HECKMAN SELECTION MODELS (BASED ON IMPUTED EX-ANTE CLOSENESS)

Heckman coefficients reported	Heckman (1)		Heckman (2)		Heckman (3)		Heckman (4)	
	(1-1)	(1-2)	(2-1)	(2-2)	(3-1)	(3-2)	(4-1)	(4-2)
Dep. Variable:	Participation	Yes-Vote	Participation	Yes-Vote	Participation	Yes-Vote	Participation	Yes-Vote
Complexity		-0.260***		-0.280***		-0.338***		-0.347***

	(1)	(2)	(3)	(4)
(proposition)	(0.053)	(0.079)	(0.082)	(0.081)
Complexity (ballot)	-0.351*** (0.106)	-0.381*** (0.121)	-0.479*** (0.131)	-0.511*** (0.119)
Exclusion Restriction				
Closeness (imputed)	-0.947*** (0.203)	-0.643 (0.481)	-0.832** (0.388)	-0.848*** (0.323)
Closeness (imputed) squared	-3.269*** (0.819)	-3.451*** (1.034)	-1.619** (0.783)	-0.643 (0.404)
Unobserved Factors				
rho	0.898	0.570	0.529	0.402
Wald test (*p-value*)	0.000	0.026	0.004	0.001
Canton dummies	No	Yes	Yes	Yes
Referenda type dummies	No	Yes	Yes	Yes
Year dummies	No	No	Yes	Yes
Policy area dummies	No	No	No	Yes
Observations	191669	191669	191669	191669

Notes: The table provides the estimated coefficients of the Heckman selection model and establishes the significantly negative effect of complexity on voter's probability to participate and vote in favor of a proposition. In contrast to Table 3, an ex-ante measure of closeness is used based on pre-poll data. Pre-poll data is available only for 47 propositions. Other values are imputed from a prediction based on Models (1) to (4) in Table A.2 panel B. Imputed values in Heckman Models (1), (2), (3), and (4) are taken from the corresponding Models (1), (2), (3), and (4) in Table A.2 panel A. The table also reports the correlation ρ and its p-value between the error terms of both equations. *Complexity* represents the log of the number of words used in the information text in the official booklets per proposition (voting equation) or aggregated at the ballot level (participation equation). All equations are estimated including individual controls as in Table 1. Heteroskedasticity-robust standard errors clustered at the ballot level are reported in parentheses. *** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level.

TABLE A.4—STATUS QUO BIAS AND THE USE OF INFORMATION BOOKLETS

	Heckman (1)		Heckman (2)		Heckman (3)	
	(1-1)	(1-2)	(2-1)	(2-2)	(3-1)	(3-2)
Dep. Variable:	Participation	Yes-Vote	Participation	Yes-Vote	Participation	Yes-Vote
Complexity (proposition)		-0.317*** (0.088)		-0.311*** (0.087)		-0.317*** (0.081)
Complexity (ballot)	-0.697*** (0.148)		-0.614*** (0.141)		-0.551*** (0.154)	
Booklet-reader			0.864*** (0.049)	0.088*** (0.028)	2.060*** (0.694)	0.022 (0.308)
Complexity (proposition) x Booklet-reader						0.009 (0.040)
Complexity (ballot) x Booklet-reader					-0.139* (0.082)	
Avg. Marginal Effects	Pr(π=1)	Pr(v=1|π=1)	Pr(π=1)	Pr(v=1|π=1)	Pr(π=1)	Pr(v=1|π=1)
Complexity (proposition)		-0.114*** (0.032)		-0.111*** (0.031)		-0.112*** (0.031)
Complexity (ballot)	-0.142*** (0.029)	0.017** (0.008)	-0.114*** (0.026)	0.011* (0.006)	-0.115*** (0.025)	0.011* (0.006)
Booklet-reader			0.177*** (0.010)	0.016** (0.007)	0.177*** (0.010)	0.016** (0.007)
Observations	128101		128101		128101	

This table establishes the robustness of the negative and significant effect of complexity on voter's probability vote in favor of a proposition when controlling for the actual use of the booklet by the voter. The variable *using information booklet* is binary and equals one if the voter reports the use of the information booklet. *Complexity* represents the log of the number of words used in the information text in the official booklets per proposition (voting equation) or aggregated at the ballot level (participation equation). All regressions are estimated with fixed effects for year, canton and referendum type and controls for individual characteristics. Heteroskedasticity-robust standard errors clustered at the ballot level are reported in parentheses. *** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level.

TABLE A.5—PERCEIVED IMPORTANCE OF PROPOSITIONS AND COMPLEXITY

Aver. Marginal Effects reported	Personal Importance				Country Importance			
	Heckman (1)		Heckman (2)		Heckman (3)		Heckman (4)	
	(1-1)	(1-2)	(2-1)	(2-2)	(3-1)	(3-2)	(4-1)	(4-2)
Dep. Variable:	Participation	Yes-Vote	Participation	Yes-Vote	Participation	Yes-Vote	Participation	Yes-Vote
Complexity		-0.140***		-0.157***		-0.137***		-0.161***
(proposition)		(0.048)		(0.047)		(0.048)		(0.047)
Complexity	-0.158***	0.026**	-0.136***	0.027**	-0.155***	0.028**	-0.155***	0.027**
(ballot)	(0.035)	(0.013)	(0.028)	(0.013)	(0.035)	(0.014)	(0.034)	(0.013)
Personal Importance			0.033***	0.018***				
			(0.002)	(0.003)				
Country Importance							0.013***	0.022***
							(0.001)	(0.003)
Observations	132022		132022		127439		127439	

Notes: This table establishes the robustness of the negative and significant effect of complexity on voter's probability vote in favor of a proposition when controlling for the perceived importance of the proposition. Both importance measures are measured on a scale from 0 to 10, where 0 = unimportant, 5 = medium importance, 10 = high importance. *Complexity* represents the log of the number of words used in the information text in the official booklets per proposition (voting equation) or aggregated at the ballot level (participation equation). All regressions are estimated with fixed effects for year, canton and referendum type and controls for individual characteristics. Heteroskedasticity-robust standard errors clustered at the ballot level are reported in parentheses. *** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level.

TABLE A.6—GOVERNMENT RECOMMENDATION AND COMPLEXITY

OLS coefficients reported	Dep. Var.: Complexity				
	(1)	(2)	(3)	(4)	(5)
Government Recommendation	-0.197	-0.163	-0.370	-0.455	-0.140
(Federal Council Advice = No)	(0.151)	(0.129)	(0.729)	(0.572)	(0.713)
Individual Controls	No	Yes	Yes	Yes	Yes
Canton dummies	No	Yes	Yes	No	Yes
Referendum type dummies	No	No	Yes	Yes	Yes
Year Dummies	No	No	Yes	Yes	Yes
Topic dummies	No	No	No	Yes	Yes
R-squared	0.006	0.102	0.109	0.217	0.329
Observations	205175	205175	205175	205175	205175

Notes: This table shows that there is no statistically significant relationship between the government's voting recommendation and our complexity measure. The dependent variable is the number of words used in the information text per proposition. OLS coefficients are reported in columns (1) to (5). The variable *government recommendation* is binary and equals one if the government recommends voting against a proposition. Heteroskedasticity-robust standard errors clustered at the ballot level are reported in parentheses. *** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level.

TABLE A.7—EXCLUSION OF PROPOSITIONS LIKELY TO SUFFER FROM SURVEY BIAS

Heckman coefficients reported	Heckman (1) exclude 5%		Heckman (2) exclude 10%		Heckman (3) exclude 25%		Heckman (4) exclude 50%	
	(1-1)	(1-2)	(2-1)	(2-2)	(3-1)	(3-2)	(4-1)	(4-2)
Dep. Variable:	Participation	Yes-Vote	Participation	Yes-Vote	Participation	Yes-Vote	Participation	Yes-Vote
Complexity (proposition)		-0.315***		-0.312***		-0.298***		-0.280***
		(0.079)		(0.078)		(0.070)		(0.074)
Complexity (ballot)	-0.536***		-0.526***		-0.523***		-0.597***	
	(0.115)		(0.117)		(0.117)		(0.113)	
Exclusion Restriction								
Closeness	-0.889***		-0.912***		-0.852***		-0.938***	

	(0.227)		(0.225)		(0.201)		(0.240)	
Closeness squared	-2.494***		-2.450***		-3.145***		-3.808***	
	(0.698)		(0.736)		(0.850)		(0.901)	
Avg. Marginal Effects	Pr(π=1)	Pr(v=1|π=1)	Pr(π=1)	Pr(v=1|π=1)	Pr(π=1)	Pr(v=1|π=1)	Pr(π=1)	Pr(v=1|π=1)
Complexity (proposition)		-0.119***		-0.118***		-0.111***		-0.104***
		(0.032)		(0.031)		(0.027)		(0.029)
Complexity (ballot)	-0.167***	0.049***	-0.164***	0.045***	-0.163***	0.043***	-0.186***	0.051***
	(0.035)	(0.014)	(0.036)	(0.013)	(0.036)	(0.011)	(0.035)	(0.016)
Unobserved Factors								
rho	0.420		0.398		0.381		0.394	
Wald test (*p-value*)	0.000		0.001		0.000		0.002	
Observations	183755		176675		154548		112335	

Notes: The table shows the robustness of the results in Table 3 w.r.t. a potential survey bias as described in Funk (2016). In columns (1) to (4), estimates for different subsamples excluding the 5%, 10%, 25%, and 50% of the propositions with the highest survey bias as reported in her appendix Table 2. Heckman coefficients are reported in the upper half of the table. Average marginal effects are reported in the lower half. *Complexity* represents the log of the number of words used in the information text in the official booklets per proposition (voting equation) or aggregated at the ballot level (participation equation). All regressions are estimated with fixed effects for year, canton and referendum type and controls for individual characteristics. Estimations that use the participation dummy as the dependent variable use complexity at the ballot level; estimations that use the yes-vote dummy as the dependent variable use complexity at the proposition level. Heteroskedasticity-robust standard errors clustered at the ballot level are reported in parentheses. *** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level.

TABLE A.8—SUBJECTIVE COMPLEXITY AS FUNCTION OF OBJECTIVE COMPLEXITY AND EDUCATION

Dep. Var.:	Difficulty to Form Opinion (ballot)				Difficulty to Form Opinion (proposition)			
Avg. Marginal Effects reported	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Complexity (ballot)	0.019**	0.092***	0.092***	0.087***				
	(0.009)	(0.023)	(0.022)	(0.018)				
Complexity (proposition)					0.049***	0.075***	0.084***	0.086***
					(0.018)	(0.014)	(0.016)	(0.013)
Education		-0.042***	-0.042***	-0.042***		-0.030***	-0.029***	-0.029***
		(0.002)	(0.002)	(0.002)		(0.002)	(0.002)	(0.002)
Individual Controls	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Canton dummies	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Referendum type dummies	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Year dummies	No	No	Yes	Yes	No	No	Yes	Yes
Policy area dummies	No	No	No	Yes	No	No	No	Yes
Observations	182639	182639	182639	182639	170094	170094	170094	170094

Notes: This table establishes the positive effect of objective complexity on subjectively perceived complexity at the ballot and the proposition level. In all specifications, education is negatively correlated with subjective complexity. The results suggest that subjective complexity is a function of objective complexity and individual characteristics such as education. Average marginal effects based on probit regressions are reported in all models. *Complexity* represents the log of the number of words used in the information text in the official booklets per proposition (voting equation) or aggregated at the ballot level (participation equation). *Difficult to Form an Opinion* represents the respondents' perceived difficulty to form an opinion about a particular proposition (voting equation) or in general with respect to all propositions on a specific ballot (participation equation). Alternative specifications using the untransformed word count produce similar results. The variable *education* ranges from 1 (mandatory school) to 6 (university). Heteroskedasticity-robust standard errors clustered at the ballot level are reported in parentheses. *** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level.