Policy area satisfaction, perceptions of responsibility, and political trust: a novel application of the REWB model to testing evaluation-based political trust

Lisanne de Bloka, Tom van der Meerb and Wouter Van der Brugb

aVrije Universiteit Amsterdam - Political Science and Public Administration, Amsterdam, Netherlands; bPolitical Science, University of Amsterdam, Amsterdam, Netherlands

ABSTRACT
Existing research demonstrates that evaluations of policies affect political trust. However, two limitations to the extant literature require us to reassess the micro-level foundation of the dominant trust-as-evaluation approach. Empirically, the common test of the evaluation-trust link relies on comparisons between individuals, an approach that suffers from endogeneity problems. Theoretically, the evaluation-trust link should be conditional on perceptions of government responsibility. Yet, this moderator is not included in most studies. We address these limitations in tandem. We study the conditionality of the evaluation-trust link on perceptions of government responsibility, using an innovative application of the Random Effects Within-Between (REWB) model, which simultaneously analyses variation between and within respondents. While our findings provide specific support for the (conditional) evaluation-trust link, we conclude that the literature has been overly optimistic when relying on between-respondent analyses.

ARTICLE HISTORY
Received 13 August 2019; Accepted 28 April 2020

Introduction
The evaluative approach to political trust argues that political trust is driven by evaluations of government performance (e.g. Huseby 2000; Kumlin 2004, 2007; A. Miller and Listhaug 1999). According to this dominant approach, people provide and withhold trust by holding governments accountable for policy outcomes, i.e. by a mechanism of punishment and reward. At the micro-level, the approach has found empirical support via the consistently strong relationship between performance evaluations and political trust. Yet, two limitations to the existing research require us to reassess this
micro-level empirical foundation of the dominant approach in the political trust literature.

First, the empirical tests of the evaluative approach commonly rely on cross-sectional studies in combination with comparisons between citizens. Such a method does not provide sufficient evidence for a punishment-reward mechanism (Van de Walle and Bouckaert 2003). The relationship suffers from endogeneity problems: those who trust the government may as a result be more positive about the policies they implemented. Moreover, correlations between political trust and performance evaluations may also be partially spurious, to the extent that both are affected by more general underlying attitudes or even response-style components. General discontentment bias holds that people who tend to be negative or pessimistic in general will simply express less trust in politicians and report lower levels of satisfaction with policies, irrespective of their evaluation of governmental policy performance (Bartels 2008; Bell and Jones 2015). The other side of the coin is that those who are generally more optimistic will tend to give more positive evaluations of public policies and will place more trust in politicians.

Here, the question is not whether current micro-level evidence for the reward-punishment mechanism is biased in favour of finding an effect, but whether this bias leads us to draw overly optimistic conclusions about the extent to which political trust is determined by policy evaluations. While several scholars have recognized that the relationship between political trust and policy evaluations may be partially spurious and/or endogenous, we are not aware of survey studies that have addressed this problem head-on.

Second, most studies overlook the theoretical necessity that the evaluation-trust link should be conditional on perceptions of responsibility. Earlier tests generally assumed that citizens hold governments responsible for positive or negative developments within a policy area (Arceneaux 2006; Arceneaux and Stein 2006; Cutler 2008; Rudolph 2003, 2006). Whether this assumption holds is, however, an empirical question. Particularly a system of multilevel government imposes great cognitive demands on citizens, as it requires information on the responsibilities, jurisdiction and past performance of each level of government. “Where there are multiple levels of government, often with different parties (or combination of parties) in office, each developing its own record, the simple punishment-reward model becomes much more complex” (Johns 2011, 54).

These twin problems cannot be tackled in isolation. To address the first, we propose a novel application of the recently established “within-between Random Effects” (REWB) framework (Bartels 2008; Bell and Jones 2015). While this framework is conventionally used to analyse multilevel and panel data, we show that it can also be used in the absence of repeated over time measurement to analyse within-person variance along a range of items. Our design provides a more crucial test of the evaluative approach
to political trust. We report results from a study that we designed explicitly for this purpose, covering performance evaluations in a range of policy areas in combination with perceptions of the responsibility of different levels of government (local, national and EU) for those policies. Moreover, we measure trust in these three levels of government. This combination allows us to assess whether people indeed place more trust in certain levels of government if they give more positive evaluations on policies for which they think this particular level of government is responsible. In order to test this, we isolate and simultaneously estimate within-respondent effects and between-respondent effects in a single model. The focus on within-respondent effects filters out potentially spurious relations between policy evaluations and political trust resulting from differences between individuals.

We address the second limitation by considering the conditionality of the punishment-reward mechanism on the perceptions of responsibility, the link that is currently missing in most existing studies (Van de Walle and Bouckaert 2003; van der Meer and Hakhverdian 2017). We argue that subjective attributions of government responsibility form an essential precondition for the reward-punishment mechanism underlying the relationship between performance evaluations and political trust (Van de Walle and Bouckaert 2003).

Theoretically, this is not entirely new in the literature on trust as an evaluation. Hobolt and Tilley (2014, 132) demonstrate that subjective attributions of responsibility to the European Union condition the effects of performance evaluations on trust in the EU. While this is an important step forward, their study falls short of identifying the precise mechanism due to data limitations. In multilevel governments, citizens are likely to compare and evaluate the responsibility and performance of government levels against each other. However, because the models of Hobolt and Tilley (2014) only include information about trust in the European Union but not trust in other levels of government, they are unable to rule out alternative explanations besides trust-as-evaluation that might drive the effects that they observe. As their analyses by necessity rely on differences between rather than within individuals, it cannot firmly distinguish between the punishment-reward mechanism and general discontentment bias as the underlying explanation for the effects. Hence, we build on and extend Hobolt and Tilley’s work by providing a comprehensive test of responsibility attribution as a necessary condition for performance-based political trust via the REWB framework.

These two improvements to the current literature enable us to employ a stringent test of the punishment-reward mechanism assumed to drive the relationship between perceived government performance and political trust. Our findings address central assumptions in two broader strands of literature. First, they specify the evaluative model that is dominant in the literature on political trust (C. J. Anderson and Guillery 1997; Dalton 2004; van der Meer and Hakhverdian 2017). While we find evidence that the reward-
punishment mechanism does indeed exist, we show that the causal relationship between performance evaluations and political trust has been overestimated in much of the existing research. Second, it provides additional evidence for the notion that the reward-punishment mechanism, dominant in studies on economic voting and democratic accountability, is conditional on perceptions of responsibility.

Theory and hypotheses

There is a burgeoning literature on the determinants of political trust (for an overview, see Zmerli and van der Meer 2017), including among other things, personality types (Freitag and Ackermann 2016; Mondak, Hayes, and Canache 2017), political socialization (Jennings, Stoker, and Bowers 2009; Mayne and Hakhverdian 2017; Norris 2011) and personal experiences (Gay 2002; Wilkes 2015). Still, as argued by Citrin and Stoker (2018, 37) in a recent review article, “[t]he most promising explanation for a change in trust is politics itself”. The literature on the determinants of political support has focused on three different types of political performance indicators.

First, the input perspective investigates whether political trust is affected by political representation, i.e. the degree to which ideological differences in the electorate are accurately represented in parliament (Fiorina and Abrams 2011; van der Meer 2017a). Here, there is an ongoing debate on whether the electorate is as heavily polarized as political elites and on the consequences of such polarization on trust and policy-making (Fiorina and Abrams 2011; Hetherington and Rudolph 2015). Second, the process perspective investigates the actions of politicians, including their trustworthiness (Whiteley et al. 2016), corruption practices (Hakhverdian and Mayne 2012; van der Meer and Hakhverdian 2017) and scandals (Chanley, Rudolph, and Rahn 2000; Hetherington and Rudolph 2008; Keele 2007). Other process factors that are also considered include the ability to take decisions and to produce “good” policies (Hibbing and Theiss-Morse 2002).

Third, the output perspective investigates the relationship between satisfaction with the content of policies and political trust (Hibbing and Theiss-Morse 1995, 2001; A. H. Miller 1974). This research tradition also includes studies that focus on policy performance, particularly on macro-economic performance (e.g. Haugsgjerd and Kumlin 2019; van Erkel and van der Meer 2016; Weatherford 1984). The central hypothesis is that people lose trust in the government when policy output is bad (e.g. the economy is in recession), whereas trust increases when output improves (e.g. a booming economy). Hence, policy evaluations feed back into political trust. It is this latter perspective – the relationship between output performance evaluations and political trust – that our paper contributes to.

Despite its longstanding theoretical status (cf. Easton 1965), policy feedback only reached mainstream political research in recent decades (Kumlin
At its core, policy feedback theory argues that “evaluation of outputs and performance may help to generate, and probably at all times will help to sustain, confidence in the authorities” (Easton 1975, 449). The majority of existing research on policy feedback focuses on one specific outcome, the influence of macroeconomic performance on political trust (e.g. Dalton 2004; Taylor 2000), possibly because the regulation of the macro economy is believed to be one of the most salient issues for citizens (Dalton 1999). Yet, “voters do not simply evaluate the economy and therefore the government, but also evaluate the government’s performance across a range of policy areas” (Marsh and Tilley 2010, 134). Nevertheless, policy outcomes in alternative domains such as health care and social insurance are rarely considered in relation to political trust (cf. Kumlin and Haugsgjerd 2017; for exceptions, see Huseby 2000; Kumlin 2007; Lühiste 2014).

Evidence that objective macro-economic performance affects political trust is mixed (van der Meer 2017b), inducing scepticism about the explanatory power of policy performance on political attitudes (Dalton 2004; McAllister 1999). By contrast, subjective measures of performance evaluations are rather consistently related to political trust (e.g. Citrin and Green 1986; Hetherington and Rudolph 2008). The direct relationship between subjective policy performance evaluations and political trust forms a very minimal requirement of the evaluative-mechanism underlying political trust. The focus on subjective measures of political performance enables “more appropriate theoretical tests of the underlying psychological processes linking perceptions to outcomes, since what individuals act upon is not necessarily what is, but what they perceive to be” (Curtis 2014, 405 emphasis in text). We build on this idea that in order to isolate the individual level mechanisms, perceptions are reality (Kumlin 2011, 114). Our first, straightforward, hypothesis is therefore:

H1: Policy performance evaluations have a positive effect on political trust.

Even though the commonly observed relationship between subjective performance evaluations and political trust forms a minimal requirement for the punishment-reward mechanism, it does not provide sufficient proof for a causal effect of policy performance satisfaction on political trust. Evidently, this relationship between perceptions of performance and political trust may also be driven by other factors resulting from heterogeneity between individuals, such as general feelings of (dis)contentment where general resentment or positivity is displaced on political objects (Kumlin 2007; Van de Walle and Bouckaert 2003). Alternatively, low levels of political trust can function as a heuristic to induce poor policy evaluations. Therefore, a more crucial test of the punishment-reward mechanism involves isolating the evaluative mechanism from general feelings of (dis)contentment. This can be achieved by
focusing on variation within individuals at one point in time (where the evaluative mechanism should hold) rather than solely on variation between individuals (where general discontentment bias is most likely).

By focusing on within-individual effects, one controls for differences between individuals, including their general predispositions towards politics, and thereby rules out many alternative explanations. The isolation of the within-person from between-person variance thus substantially increases the internal validity of causal claims.

The increased complexity of multilevel democracies makes evaluating government performance more cumbersome, since citizens need to be aware of the performance and responsibilities of each level simultaneously, even though they strongly vary in size, visibility, capacity and jurisdiction. These information costs make it harder for citizens to allocate clear responsibility for outcomes to the correct level of government (C. D. Anderson 2006; Cutler 2004, 2008; Hobolt and Tilley 2014).

Perceptions of responsibility for policy outcomes are the key moderator that ought to condition the relationship between evaluations of policy area satisfaction and political trust (Easton 1975; Tyler 1982). Hobolt and Tilley (2014) convincingly show that perceived EU responsibility conditions the relationship between economic and health care evaluations and trust in the EU. Similarly, in the economic voting literature a handful of studies have incorporated subjective measures of responsibility into their models (Marsh and Tilley 2010; Rudolph and Grant 2002). Irrespective of the actual division of responsibilities between different levels of government, the reward-punishment mechanism will only operate when authorities are perceived to be responsible. That is not to say that the objective distribution of government power is irrelevant, as these structures may affect the degree of perceived responsibility. Our second hypothesis is therefore:

H2. The effects of policy performance evaluations on political trust are moderated by perceptions of government responsibility: When a governing body is perceived to be more responsible for policy outcomes, the effect of policy performance evaluations increase.

To the extent that political trust is based on a cognitive evaluation, conditional on one’s perception of responsibility, we expect to see this moderation effect within respondents. Such a moderation is much less likely to play a role between persons, as between-person variation deals with the relationship between general performance evaluations and general political trust. While

---

1Perceptions of responsibility can be affected by predispositions about a level of government (group-serving biases). There is some support for selective attributions of responsibility along partisan lines at the national level (Hobolt, Tilley, and Wittrock 2013; Tilley and Hobolt 2011). Yet, this need not affect these attributions in a multilevel setup. Our correlation matrices (Table 3, Appendix A) show that the perceptions of responsibility are not or only very weakly ($r = 0.15$) related to levels of political trust.
it is the most studied, this between-person relationship is not necessarily conditional on the general level of perceived responsibility of all levels of government due to alternative causal explanations, such as general discontentment bias, that are unaffected by perceptions of responsibility.

When testing H2, we will allow for the possibility of a non-linear relationship. It seems plausible that responsibility perceptions exert a weak effect on the evaluation-trust relationship in policy areas where governments are thought to be hardly responsible anyway or in areas where people do not have very outspoken ideas about government responsibilities.2

Data and methods

To test the hypotheses put forward in this paper, we employ data from the first Dutch Local Election Study (DLES). The DLES consists of a single wave of interviews among a random subsample of the Longitudinal Internet Studies for the Social sciences (LISS) panel. The LISS-panel is a high quality randomly-selected probability sample of all Dutch households drawn from the population register by Statistics Netherlands. The complete panel consists of 4500 households, containing 7000 individuals, and is refreshed every 8 years.3

The DLES drew a sample of respondents from the participants in the overarching panel. This sample was restricted to panel members of eighteen years and older. Fieldwork for the DLES (by means of an online questionnaire) was completed in March 2016 and resulted in a response rate of 79.6% of the approached members of the LISS panel (n = 2,579). While the data were collected by interviewing members of a panel, our research design is cross-sectional.4

Dependent variable

Political trust is measured for three levels of government: the local level (municipality), the national level (country), and the supranational level (European Union). For each of these three levels, several survey questions were posed to measure our dependent variable: satisfaction with the way democracy

2Seeing that multilevel systems pose such great cognitive demands on citizens, it is possible that the degree to which responsibility perceptions play a role for the performance-trust link between citizens is influenced by the level of political knowledge. We have tested this using a three-way interaction effect of knowledge, perceived responsibility and performance satisfaction using a subsample of our data (see Appendix B). We find that, for most policy issues, there is indeed a positive significant three-way interaction effect: Among those with higher levels of political knowledge, responsibility attributions more strongly moderate the relationship between policy area satisfaction and political trust. In the scope of this article, we decided not to pursue this path. Nonetheless, it does suggest that there could be an evaluative mechanism driving political trust at the between level, under the circumstances that political knowledge is high.

3Panel members are provided a computer and internet connection if they otherwise could not participate and are paid a small amount for each completed survey.

4As a pilot study, the DLES 2016 was not held at the time of any election.
works (SWD) and trust in core political institutions (Dalton 2004; Norris 2011). SWD at three government levels is measured on a four-point scale, ranging from 0 (“completely dissatisfied”) to 3 (“very satisfied”). Trust in the institutions of each government level is based on a series of questions asking how much the respondent trusts an institution on a four-point scale ranging from “completely no trust” to “a lot of trust”.

Trust in the national government consists of one’s trust in the Government and the Lower House of Parliament (Second Chamber) ($\alpha = 0.81$). Municipal level trust is made up of one’s trust in the Municipal Council, the Local Executive, and the Mayor ($\alpha = 0.90$). European level trust is comprised of one’s trust in the European Union.

SWD and trust in government institutions were combined into scales of political trust by summing the scores on the indicators and dividing this by the number of items (Cronbach’s $\alpha$ is 0.72 for the municipal level, 0.75 for national level, and 0.84 for the European level). For each level, we thus created a scale that ranges from 0 to 3. In line with the literature (cf. Muñoz 2017), political trust is generally highest for the municipal level (M: 1.70), followed by the national level (M: 1.59), and the European level (M: 1.06). Our dependent variables of political trust contain 7 (EU), 13 (national), and 30 (local) unique values between 0 and 3 and is treated as a continuous variable.

**Independent variables**

**Perceived responsibility**

Respondents indicated how responsible they thought each level of government was for policy in six domains: crime prevention, health care, social security, pensions, railway services, and refugees. Perceived responsibility is measured on an 11-point scale, ranging from 0 (not at all responsible) to 10 (highly responsible). The six policy domains had been selected to ensure sufficient variation between different levels of government in actual responsibility. Furthermore, the questions were formulated carefully to avoid ideological positioning.

Figure 1 shows that perceived government responsibility per level varies per policy area. The national government is deemed most responsible on average for all policy domains, even those that have been privatized (such as railway services) or decentralized (such as youth and elderly care). The local government is perceived as particularly responsible for crime prevention and care for the youth and elderly. The EU is perceived as relatively responsible for crime prevention and asylum-seeking refugees.

---

5The option “don’t know” is coded as missing.
6The outcomes are robust to the use of single items rather than index variables (see Appendix C).
7To limit ideological bias, these measures do not cover policies but policy areas. For the full question wording, see Appendix D.
8In Appendix E we tested the robustness of our models to straight-lining (7.3% of our sample), finding similar effect sizes.
Figure 1. Mean responsibility scores per area. Source: DLES 2016.
**Policy area satisfaction**

To measure policy area satisfaction, respondents were asked to evaluate the current state of the selected policy areas. Answers range from 0 “very dissatisfied” to 10 “very satisfied”. Don’t know is coded as missing (11.64%). Respondents were most satisfied with social security (M: 5.27), railway services (M: 5.51), and crime prevention (M: 5.50) and least satisfied with the current policy on asylum seeking refugees (M: 4.13).

**Control variables**

We control for various rival explanations of political trust that are put forward in the literature: age, level of education, gender, political interest, political knowledge, left-right ideology, and support for a government party (according to the reported vote in the previous elections) (see Appendix F for descriptive statistics). These control variables are only included at the between-level (see below), as they only vary between respondents.

**Modelling technique**

To deal with the two limitations in the literature on the evaluation-trust relationship we identified above, our modelling strategy needs to (i) estimate the conditionality of that relationship, and simultaneously (ii) separate within- and between-person variance in a single model. In isolation, either of these strategies would be insufficient for a more crucial test.

We transposed our data into a long format with each respondent having 18 observations (all combinations of three levels of government and six policy areas). On this dataset we perform a multilevel linear regression (using the “xtreg” command in Stata 13) including both between- and within-effects of policy area satisfaction and perceived responsibility (level 1 variables). The novel “within-between Random Effects” (REWB) framework for analysing nested/panel data (Bartels 2008; Bell and Jones 2015) allows us to separate these effects. The REWB framework was designed to analyse different types of clustered data, including panel, time-series cross-sectional (TSCS), and multilevel data (Bartels 2008, 1) without inflating the effective sample size or artificially suppressing standard errors. We show that meaningful separation of between and within-person variances is also possible with micro-level cross-sectional data, when one is dealing with multiple measurements. These multiple measurements need not be temporal or geographical, but can also entail

---

9As a robustness check, we have also tested our hypotheses using a dataset in which government levels were clustered within respondents (3 per respondent). Here, responsibility perceptions and political trust vary across government areas. We ran separate models for each policy area (see Appendix G). The findings are very much in line with those that we find on the combined dataset (with 18 observations within individuals). Although there is a strong between-person effect of performance satisfaction on political trust, this is not moderated by perceived responsibility. Rather, the moderation between responsibility perceptions and performance satisfaction only occurs within individuals.

10Figures are created using the “plotplain” graphic scheme (Bischof 2017).
the measurement of a single construct over various topics (e.g. evaluation of different policies, trust in various government levels, or vote propensity scores) (for a similar application, see Werner 2020).

By transposing the dataset, we create a multilevel dataset with observations clustered within respondents. We thereby model the unobserved cluster heterogeneity that may affect our results. As with the original application of the REWB model, we are now dealing with multiple measurements of a single construct (of trust, policy evaluation and perceived responsibility) within one individual. The REWB framework has the advantages of fixed effects and random effects models combined (Bartels 2008; Bell and Jones 2015). Rather than controlling for cluster confounding bias between higher level units (i.e. respondents), as is done in fixed effects models, the REWB framework explicitly models such (causal) heterogeneity by separating between-from within-cluster effects. This “[c]luster confounding has significant implications for how one interprets the effects of independent variables in clustered data, and therefore, detecting and correcting for it is crucial for understanding the precise nature of relationships and for testing hypotheses” (Bartels 2008, 10).

The between-respondent effect of responsibility attribution constitutes the average tendency of a respondent to assign responsibility to policy areas to a level of government, while our within-respondent effect of responsibility attributions specifically grasps the variations from this mean in the respondent’s assignment of responsibility for various issues to the local, national and EU government. The between-respondent effect of policy area satisfaction constitutes the mean evaluation across all issues, while the within-effect entails the variations from this mean, i.e. when satisfaction with one policy area is higher in relation to the others. This separation satisfies the statistical assumption that level-1 independent variables should be uncorrelated with the random effects term.11

This use of the REWB framework helps to isolate the endogeneity effects we discussed above, which primarily affects between-respondent effects. Yet, it does not fully eliminate endogeneity problems. Specifically, predispositions such as trust in government might function as a heuristic to judge the performance of that government or to selectively attribute responsibility. We argue, however, that this does not pose serious threats to our analyses for two reasons. First, endogenous effects can only offer an alternative to the theoretical model on the conditionality of the performance-trust link we proposed above if respondents’ policy area satisfaction and trust levels were set, and responsibility attributions were used to harmonize the two. Yet, this does not align with the literature on the partisan bias, which implies that causality runs from political trust or partisanship to performance.

---

11For the full specification of the models, see Appendix H.
evaluation rather than both of these being predetermined (e.g. Bartels 2002). Second, we limited the risk of reversed causality by posing the survey questions on evaluation after those on perceived responsibility.

Results

Main effects

Model A (Table 1) tests the first hypothesis on the main effect of policy area satisfaction on political trust. The model presents the results from a linear random intercept model including the main effect of policy area satisfaction. It distinguishes between within-individual effects and between-individual effects, and includes the absolute value of the difference between the within and between-individual effects to test for cluster confounding (cf. Bartels 2008, 18). The $\rho$ suggests that a more than half (56%) of the remaining error variance is accounted for by the between-respondent level error.

The results indicate that the between-respondent variable for policy area satisfaction has a positive significant main effect on political trust: Individuals with higher levels of policy area satisfaction in general have higher levels of overall trust for governments compared to those with more negative evaluation of governments. This is a finding similar to that of previous research. However, as we argued above, this finding does not constitute sufficient evidence for the evaluative mechanism, as other causal mechanisms may drive the effect, including a more basic positivity/negativity bias that drives both evaluations and trust.

The direct within-effect of policy area satisfaction on political trust is positive but very small and statistically not significant. This was to be expected, given the fact that, at the within respondent level, policy area satisfaction does not vary across government levels. The abs. test shows that there is significant cluster confounding for policy area satisfaction: the between- and within-effect of satisfaction on political trust are significantly different. H1 – on the main effect of general evaluations of policies on general political trust – thus only finds policy area satisfaction as a significant between-respondents and not a within-respondents effect.14

12For the complete table, see Appendix I.
13Our policy area satisfaction questions are measured for different issues, rather than for different levels of government. This was a conscious choice, because we wanted respondents to evaluate policy domains without having a level of government in mind (due to the endogeneity problems mentioned in the paper). We therefore only have variation across six observations out of the 18 within individuals. In addition, since respondents only vary moderately in their policy area satisfaction, there is relatively little variation within-individuals. As such, it is rather unlikely to find direct within-respondent effects of policy area satisfaction on political trust.
14We performed various robustness checks which led to similar findings: (1) a within-between RE model with two separate dependent variables, (2) a fixed effects model with two separate dependent variables using OLS. See Appendices C and J.
Table 1. Direct & conditional effects on political trust (random-effects maximum likelihood estimation model).

<table>
<thead>
<tr>
<th>Variables</th>
<th>Model A: Main effects</th>
<th>Model B: Conditional effects</th>
<th>Model C: Non-linear effect</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Between-effects</td>
<td>Within-effects</td>
<td>Abs test</td>
</tr>
<tr>
<td>Policy area satisfaction</td>
<td>0.59***</td>
<td>0.00</td>
<td>0.59***</td>
</tr>
<tr>
<td></td>
<td>(0.02)</td>
<td>(0.00)</td>
<td>(0.02)</td>
</tr>
<tr>
<td>Perceived responsibility</td>
<td>0.05*</td>
<td>0.05***</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td>(0.2)</td>
<td>(0.00)</td>
<td>(0.02)</td>
</tr>
<tr>
<td>Perceived responsibility (squared)</td>
<td>0.05*</td>
<td>0.05***</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td>(0.12)</td>
<td>(0.01)</td>
<td>(0.02)</td>
</tr>
<tr>
<td>Dummy DK responsibility</td>
<td>0.01</td>
<td>−0.00</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td>(0.02)</td>
<td>(0.01)</td>
<td>(0.02)</td>
</tr>
<tr>
<td>Satisfaction*Perceived Responsibility</td>
<td>−0.11</td>
<td>0.09***</td>
<td>−0.21*</td>
</tr>
<tr>
<td></td>
<td>(0.09)</td>
<td>(0.02)</td>
<td>(0.09)</td>
</tr>
<tr>
<td>Satisfaction*Perceived responsibility (squared)</td>
<td>−0.77***</td>
<td>0.10**</td>
<td>−0.92***</td>
</tr>
<tr>
<td></td>
<td>(0.35)</td>
<td>(0.04)</td>
<td>(0.35)</td>
</tr>
<tr>
<td>Satisfaction*Dummy DK</td>
<td>−0.04</td>
<td>−0.03</td>
<td>−0.00</td>
</tr>
<tr>
<td></td>
<td>(0.09)</td>
<td>(0.03)</td>
<td>(0.09)</td>
</tr>
<tr>
<td>Sigma u</td>
<td>0.13</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sigma e</td>
<td>0.12</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ρ</td>
<td>0.55</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LR Test</td>
<td>$\chi^2$ 21,052.38 ***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AIC</td>
<td>−51,173.91</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BIC</td>
<td>−50,967.29</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Source: DLES 2016. $N = 2,460$, $T$ (avg. 16.5); Total number of observations = 40,489.

For full table including controls, see Appendix I. Tables show unstandardized beta coefficients; *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$. 
Interestingly, model A shows that perceived responsibility also has a positive and significant main effect on political trust, between- as well as within-individuals.\textsuperscript{15} We did not theorize or anticipate this effect. For a given individual, higher levels of perceived responsibility result into higher levels of trust. And more responsibility assigned to a specific level on a specific policy area results into more trust for this level. We can only speculate about the underlying causal mechanism.

\textit{Conditional effects}

Testing hypothesis 2, Model B (Table 1) shows that the direct between-respondent effect of policy area satisfaction on political trust is not conditional on perceptions of government responsibility. As expected, the between-respondent interaction effect of policy area satisfaction and responsibility perception is not statistically significant. In other words, the relationship between a respondent’s mean level of policy area satisfaction and his/her mean level of political trust is not conditioned by his/her mean level of perceived government responsibility. This provides further evidence that between-respondent effects of evaluations on trust – the evidence that is most common in this literature – are not necessarily driven by an evaluative-mechanism and may be spurious or endogenous.

Interestingly, the within-respondent interaction effect of policy area satisfaction and responsibility perception is statistically significant and positive, even when we model it as a linear effect. The cluster confounding test shows that the within-respondent effect is significantly different from the between-respondent effect. While Model A showed no signs of a main effect of policy area satisfaction on political trust at the within-level, Model B shows that a positive marginal within-effect of policy area satisfaction exists under the specific conditions of high perceived responsibility. Policy area satisfaction is more strongly related to political trust for government level X by subject S if subject S perceives government level X to be more responsible for that policy area. Figure 2 facilitates the interpretation of this result by visualizing the increase in effect size of the within-respondent policy area satisfaction when perceived responsibility increases. It illustrates that evaluations of policy areas have a stronger effect on political trust at a specific level of government when that level of government is perceived to have more responsibility for those policies.

Both the insignificant between-respondent and the significant within-respondent interaction effects are consistent with our

\textsuperscript{15}The effects are almost equal in size and the cluster confounding test shows the differences between the two are insignificant. Thus, the effect of perceived responsibility can be seen as a pooled estimate of which the interpretation does not rely on the comparison that is being made.
predictions. The more precise test of the evaluative model lies within respondents who evaluate policy areas, assign responsibility to various government levels, and allocate political trust accordingly. This supports hypothesis 2a.

**Non-linear conditional effects**

In our theory section we argued that there are theoretical reasons to expect the conditional effect to be non-linear. In model C we therefore tested to what extent the conditionality of performance-based trust on perceptions of responsibility constitutes a non-linear relationship. Figure 3 visualizes this relationship. This pattern suggests a threshold-effect: only when the respondent assigns relatively more responsibility to government level X for policy area A than to other levels of government, satisfaction with policy area A influences trust in government level X. This relationship increases exponentially as perceived responsibility increases. When government level X is considered only averagely responsible for policy field A, there is no significant marginal effect. In sum, the conditionality of the evaluation-trust link on perceptions of responsibility only takes place once perceived responsibility passes a certain threshold. There is only a strong and positive conditionality once a respondent assigns an above-average level of responsibility to a level of government.16

![Figure 2](image.png)
Discussion and conclusion

A classic punishment-reward model of political trust assumes that citizens’ confidence in political authorities depends largely on their evaluation of public policies. However, over the past decades, Western societies have witnessed a substantial erosion of the authority of national governments. Various responsibilities and jurisdictions that once belonged to the nation state have been moved to local levels, the supranational level (the EU), or even outside the political realm as a result of privatization. The complexity of multilevel governments imposes great cognitive demands on citizens: It requires information on the responsibilities and jurisdiction of each level of government. As a result, it has become increasingly complex for citizens to hold governments accountable for their performance.

On the one hand, this makes the conventional reward-punishment model of political trust more complex: the relationship between policy area satisfaction and political trust should be conditional on perceptions of responsibility. On the other hand, this very same complexity enables a stricter test of the

---

Figure 3. Marginal within-respondent effect policy area satisfaction on political trust, conditional on perceived responsibility, including 95% confidence intervals. Source: DLES 2016.

---

While it would certainly be interesting to see how the conditionally works for specific levels of government, or for specific policy areas, the current research design does not allow for second-order interactions with either policy areas and/or levels of government. Because the data is transposed into a long format on the basis of policy area and government level combination, including such a second-order interaction term would result in a lack of variance either on the dependent variable (when interacted with level of government) or on the independent variable (when interacted with policy area).
trust-as-evaluation approach against problems of endogeneity such as a general discontentment bias.

This study presents a novel application of the “within-between Random Effects” (REWBe) framework, that models the variances within-persons (across policy fields and levels of government) and between-persons (irrespective of either) simultaneously. Our analyses show that the relationship between policy area satisfaction and political trust is much stronger at the between-person level than at the within-person level. However, the former does not provide strong evidence of a causal effect as the relationship may be partially spurious. Our finding that evaluations of responsibility do not condition the relationship between policy area satisfaction and political trust at the between level, suggests that the relationship is indeed to a large degree spurious. After all, if policy area satisfaction would be the driver of trust in governments, we would expect this to be mostly the case for those issues for which these governments are seen to be responsible (Hobolt and Tilley 2014). The likely spuriousness of this relationship at the between respondent level is an important finding, as comparisons between respondents form most of the existing evidence of the evaluation-trust relationship. In other words, due to endogeneity and spuriousness, extant micro-level research tends to seriously overestimate support for the evaluation-trust relationship when it relies on between-person comparisons.

Yet, concurrently, our findings do offer clear support for the reward-punishment mechanism underlying political trust. Attributions of responsibility are indeed a crucial precondition for this mechanism. The higher the attributed responsibility to a level of government, the stronger the marginal effects of policy area satisfaction on political trust. Our findings thus confirm the conclusions of Hobolt and Tilley (2014), who showed that the effect of policy evaluations on trust in the European Union is conditional on the extent to which respondents perceive the EU to be responsible for those issues. The data employed by Hobolt and Tilley allowed them to study this interaction only at the between-respondent level. The fact that we find the same conditional effect at the within-respondent effect (comparing between six issues and across three levels of government), lends stronger credence to the causal evaluation-trust mechanism. The occurrence of this moderation effect as a within-respondents rather than a between-respondents mechanism provides important evidence.

Theoretically, our findings stress the importance of responsibility attributions. Yet, rivalling explanations on the determinants of political trust are not mutually exclusive. The evaluative approach finds credence in the within-person effects: citizens do evaluate policy performance and hold governments accountable for policy outputs. Concurrently, socio-cultural determinants of political trust may help explain why some people hold higher
levels of trust in general, irrespective of whom their perceive to be responsible for specific positive or negative developments (the between-effect).

Methodologically, these findings underpin the relevance of separating within- from between-person effect, as we did in the novel application of the unified within-between Random Effects framework (Bartels 2008; Bell and Jones 2015). We illustrate that this versatile framework is not only applicable to panel and multi-level data, but is also a powerful tool for within-respondent variance along multiple measures in many fields in political sociology and political psychology. The separation of within- from between-person variance provides the necessary leverage to distinguish between rivaling explanations of the evaluation-trust relationship.

Our approach does not provide a panacea to all endogeneity problems discussed in the literature when studying survey data and investigating political trust and policy area satisfaction in particular. For example, the monolith measures on perceived policy area satisfaction do not allow us to differentiate between policy areas or between each government level within the REWB framework. As such, the next step to test the evaluative-trust mechanism requires an extension of the current within-between respondents model with longitudinal data on changes in policy area satisfaction, responsibility attributions, and political trust. A focus on within-person changes over time will allow us to further isolate the mechanisms at play. Here, it would particularly be interesting to investigate heterogeneity in the accountability mechanisms for different levels of governments and between policy areas.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This research was supported by the Netherlands Organisation for Scientific Research (NWO) [Nederlandse Organisatie voor Wetenschappelijk Onderzoek] under the NWO TOP grant (407-13-029).

Notes on contributors

Lisanne de Blok is postdoctoral researcher at the Vrije Universiteit van Amsterdam working on democratic legitimacy in the EU.

Tom van der Meer is professor in Political Science, in particular Legitimacy, Inequality and Citizenship at the University of Amsterdam.

Wouter van der Brug is also a professor of Political Science at the University of Amsterdam. His research interests focus on comparative research in collective political behaviour.
References


