

VU Research Portal

Philanthropy in the welfare state

de Wit, A.

2018

document version

Publisher's PDF, also known as Version of record

[Link to publication in VU Research Portal](#)

citation for published version (APA)

de Wit, A. (2018). *Philanthropy in the welfare state: Why charitable donations do not simply substitute government support*. [PhD-Thesis - Research and graduation internal, Vrije Universiteit Amsterdam].

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

E-mail address:

vuresearchportal.ub@vu.nl

While Chapter 1 gave important insights in the variation of government support and charitable donations across countries, its cross-sectional design did not provide much evidence about the causal relationship. Previous studies aimed to address causality in different ways. Therefore, the current chapter systematically reviews previous studies on the crowding-out hypothesis. It finds that about two-thirds of previous estimates find a negative correlation (crowding-out), while one third of the estimates find a positive correlation (crowding-in). The results are strongly shaped by the research methods that are used. In experiments, a \$1 increase in government support is associated with an average \$0.64 decrease in private donations, while non-experimental data analyses find an average increase of \$0.06. Random-effects regression models show that, contrary to arguments that are prevalent in the literature, studies that take subsidies to organizations as a measure of government support are less likely to estimate crowding-out than studies that use a measure of direct government expenditures. Central government support is associated with higher charitable donations, while measures that include multiple levels of government tend to find negative correlations. The results challenge the consistency of prior research findings and demonstrate the contextual dependence of the validity of the crowding-out hypothesis.

Chapter 2

A meta-analysis of the crowding-out hypothesis

Arjen de Wit and René Bekkers

RB designed the study; AdW collected the data and carried out data analysis; AdW and RB contributed to writing the article.

This chapter is published as: De Wit, A., & Bekkers, R. (2017). Government Support and Charitable Donations: A Meta-Analysis of the Crowding-out Hypothesis. *Journal of Public Administration Research and Theory*, 27(2), 301-319.

Data, syntax and supplementary materials are available through the Open Science Framework at <http://osf.io/6jwh2/>.

INTRODUCTION

How does the level of fundraising income of nonprofit organizations respond to changes in government funding? Over the last few years, nonprofit revenues in Western democracies have been pressured due to the economic downturn and unreliable government funding. At the same time, government policies both in the US and abroad seek to increase the role of profit and nonprofit actors in the private sector. Forms of governance that received a lot of attention include the outsourcing of public services through contracting (Smith & Lipsky, 1993), the involvement of non-state actors in consensus-based decision making (Ansell & Gash, 2008) and the emergence of interorganizational networks to deliver public services (Milward & Provan, 2003). There has been much debate about the effectiveness of different government-nonprofit collaborations. Besides internal characteristics like the institutional structure and management styles, an important condition for effective collaborations is the availability of resources in the organizational context (Ansell & Gash, 2008; Milward & Provan, 2003; Pfeffer & Salancik, 1978). Public goals can be funded through government support in the form of expenditures, subsidies, contracts or tax incentives, but also through nonprofit fundraising income. Despite the large body of governance literature, it is still unsure how different funding streams interact.

There is a wide array of studies dedicated to the crowding-out hypothesis, which claims that increasing government contributions, financed through taxes, are associated with reducing charitable donations from private donors. In earlier literature reviews Steinberg (1985, 1997) concludes that there is evidence for partial crowding-out. Payne (2009) discusses how different studies find different results and concludes that “crowdout exists—at least sometimes” (Payne 2009: 181). From a sample of 46 published and unpublished non-experimental studies, Tinkelman (2010: 24) concludes that “the results vary tremendously” and that the effect of government support depends on a number of assumptions, like full information and the costs of providing public goods. The variety of findings raises the question which conditions influence the estimated relationship between government funding and private contributions.

The current meta-analysis examines estimations of crowding-out as well as methodological and contextual characteristics in previous empirical articles. This contributes to the crowding-out literature in two ways. First, mapping methodological differences is extremely useful for further research in

this area. A better understanding of the consequences of different methodologies allows for a sensible comparison between previous results and more careful future research design choices. Second, mapping contextual differences yields theoretically useful insights on the conditions under which high government support is associated with lower charitable donations. Our meta-analysis builds upon earlier literature reviews (Steinberg, 1985, 1997; Payne, 2009; Tinkelman, 2010) by mapping differences between empirical findings in a more systematic way, providing robust evidence on contextual characteristics that are often hypothesized to be moderating variables but never tested as such. A meta-analysis is suitable for testing the conditions under which a relationship occurs. However, although we test a variety of possible moderators, there are many other theoretically relevant conditions that we are not able to test here.

Both public and nonprofit managers benefit from robust information about the effects of different types of government funding. Policy makers need to know how policy programs can be funded through effective public-private networks. Evidence that high levels of public funding are detrimental for charitable giving would support ideas about government programs with small roles for public actors and large roles for nonprofit organizations that are dependent on private funding. From the side of nonprofits, it is important to know how revenue streams interact. Organizations that heavily rely on government subsidies are likely to have a lower organizational autonomy (Froelich, 1999; O'Regan & Oster, 2002; Verschuere & De Corte, 2014). More refined knowledge about the effects of public funding on fundraising income would enable nonprofit managers to better position their organizations between government, local communities and other private actors.

The outline of this paper is as follows. In the next section we present hypotheses on the correlates of crowding-out estimates in previous research. In the Data and Methods section we present the methodology of the meta-analysis, while the Results section contains Analyses of Variance (ANOVA) and multivariate regression models to show how different study characteristics are correlated with the direction and magnitude of crowding-out that is estimated. The article closes with a discussion and conclusion.

THEORY AND HYPOTHESES

In this section we formulate hypotheses on the correlates between characteristics in research design and the crowding-out estimate. We distinguish between hypotheses on data source, sample country, regression model and specification, and operationalization of the independent variable.

Data source

Four types of data are used to test the relation between government support and charitable donations: laboratory experiments, survey experiments, archival (financial information) data and micro-level survey data. Lab experiments differ from real-world settings in “the nature and extent of scrutiny, the emphasis on the process by which decisions are made, the artificial limits placed on the action space, the imposition of task, the selection rules into the environments, and the stakes typically at risk” (Levitt & List, 2007: 168). However, the defining characteristics of laboratory experiments do not necessarily bias their outcomes in a systematically positive or negative direction. Camerer (2015) argues that laboratory and field experiments often find the same results and that the problems with generalizability of lab experiments are exaggerated.

In the case of donors’ reactions on government support, we hypothesize that laboratory experiments create a controlled environment with settings that make it more likely for crowding-out to occur. First, participants typically receive full information on the behavior of the “government” as simulated by the researchers. Most of the crowding-out experiments have a repeated-measure design in which participants not only are aware of the level of government support but also of changes therein, making it more likely that they change their giving behavior in different treatments. Horne et al. (2005) show that in reality many donors do not know how much public subsidies organizations receive. Second, participants are more sensitive to social cues because they know that they take part in a study. If people see changes in government support they suspect that this is supposed to affect their giving and, aware of being watched, they will change their donations. Especially people whose preferences are supportive of private donations as a substitution for public expenditures may be sensitive for such information. Third, participants in crowding-out experiments are almost always undergraduate students, arguably non-representative samples scoring lower on different measures of prosocial behavior and being more responsive to experimental

manipulations (Henrich, Heine, & Norenzayan, 2010). Carpenter, Connolly, and Myers (2008) show that students, especially males, give considerably lower amounts in experiments than a sample drawn from a broader population, but other studies showed that students and non-students do not differ in their level of giving in a dictator game with charities as recipients (Bekkers, 2007) and in their change in giving as a reaction to changes in other participants' donations in trust games (Falk, Meier, & Zehnder, 2013). Fourth, participants in experiments receive an endowment from the researchers, making it easier to change levels of giving than in situations where they decide on their own expenditures. Although the relative financial impacts in experimental conditions might be large (e.g. a 25% tax on a \$20 endowment), it is easier to spend money that you have not yet earned.

At least the first two characteristics of lab experiments also hold for survey experiments. A survey experiment is a randomized control trial that is part of a questionnaire, in which respondents receive different questions or pieces of information. In contrast to lab experiments, survey experiments are often carried out among a sample that is representative of the population. The only published survey experiment on crowding-out that we know of is a vignette experiment without any earnings for the participants (Kim & Van Ryzin, 2014).

Crowding-out can also be tested with archival data, for example when adopted from the U.S. Internal Revenue Service 990 financial information forms. Despite serious doubts about the accuracy of reported information on 990 forms, data that organizations report in these forms are highly correlated with those in audited financial statements (Froelich, Knoepfle, & Pollak, 2000). Organizations' income from private donors is a relatively valid measure of aggregate real-world charitable donations.

As a final data source, crowding-out studies can use survey data on individual donations that are paired with financial data on government support from other sources. Although survey research has its own issues like sample selectivity and social desirability, self-reported micro data approximates donations that are made in absence of the conditions in experimental designs.

A major concern for empirical crowding-out research is endogeneity. As described by Payne (2009), government support and private donations may be jointly determined by unobserved variables. Voter preferences for public goods might drive donations to these goods as well as government funding through the political process, generating an upward bias in the association between government support. Also, places where the need is more urgent

(e.g. high poverty) are likely to receive both high levels of public and private funding. Finally, a predictor of nonprofit revenues might be previous government grants as organizations have growing success over time (Foster & Fine, 2007). Omitted variables such as voter preferences, (changes in) the need for public goods and previous government funding upwardly bias the relation between government support and charitable donations, which would result in estimations of (in the case of crowding-out) a less strongly negative or (in the case of crowding-in) a more strongly positive association between government support and charitable donations. These concerns apply to studies using archival or survey data. Well-designed experiments are not affected because the treatment (i.e., the level of government support) is randomly assigned and participants generally cannot affect levels of government support (exceptions are experiments that allow voting, such as Blanco, Lopez, & Coleman, 2012; Isaac & Norton, 2013; Sutter & Weck-Hannemann, 2004).

In both archival and survey data, donors do not necessarily receive information on the actions of the government, the researcher demand effects are absent or weaker, the samples are generally less selective and participants report on decisions about their own (rather than the experimenters') money. Experiments are able to measure the relation between two variables in a controlled environment, while studies using financial data from surveys or archives have to deal with other factors that interfere.

H1: Studies using experimental data are more likely to find crowding-out and find stronger crowding-out than studies using non-experimental data.

Sample country

Almost all published crowding-out studies come from Western countries, which is an important caveat of the literature since the effect of government policies might be different in developing countries. But even among Western countries people and organizations may react in systematically different ways to changing policies. People from different countries differ in their stance towards social problems as requiring action from private citizens and charitable organizations or government intervention. Citizens in different countries show systematically different levels of support for extensive provision of public services by the government (Andress & Heien, 2001; Svallfors, 1997). People who are used to extensive welfare state arrangements expect the government to take care of public services and might be reluctant

to compensate for changing levels of government provision of public goods. In countries where public services are considered a shared responsibility for public and private actors, on the other hand, donors might be more willing to raise the level of donations to nonprofit organizations in order to reach the desired goals.

A possible explanation for country differences is that the marginal utility of donations decreases with the extensiveness of welfare state programs. It has been argued that the marginal increase in well-being derived from income is high for poor countries but diminishes with economic prosperity (Inglehart, 2000). The need for public or private provision of public services is more urgent in countries with more severe social problems. Welfare states differ in size and inclusiveness, and thus in their efficacy when aiming to alleviate problems like poverty, hunger and homelessness. Given that social needs are higher in countries with smaller welfare states, an additional dollar of contributions to alleviate those needs has a higher value for recipients compared to countries with extensive welfare states and less urgent social needs. It is likely that donors are more inclined to compensate for changing government support when the stakes are higher.

Also, the nature of collaborations between governmental and nonprofit actors is different in different countries. Discussing the development of “governance regimes” in Western Europe, Bode (2006: 355) perceives “a growing distance between voluntary agencies and both the welfare state and civil society; with more volatile public–private partnerships; and with a dispersed involvement of volunteers and donors.” Smaller government involvement may cause more volatile nonprofit management with a stronger focus on fundraising (Froelich, 1999; O’Regan & Oster, 2002), so organizations should be better able to respond to changing government policies.

In sum, people in countries with smaller welfare states, where the needs are more urgent, public goods are less strongly perceived as government responsibility and nonprofit management is more volatile, should be more likely to compensate government support than countries with extensive government arrangements.

H2: Studies are more likely to find crowding-out and find stronger crowding-out in less generous welfare states than in more generous welfare states.

Regression model and specification

As explained above, studies using non-experimental data are most likely to suffer from endogeneity bias. A first issue is omitted variable bias. Causal claims that are inferred from regression analyses rely on what Angrist and Pischke (2009) call the conditional independence assumption, also known as selection on observables, meaning that only observed variables account for the correlation between the independent variable and the error term. When regressing levels of charitable donations on levels of government support, this assumption is unlikely to be met, since omitted variables such as the need for public goods might upwardly bias the estimated relationship between government support and charitable donations. A related issue is the endogeneity that occurs when both the independent and dependent variable are jointly determined. If government policies reflect the same political preferences that underlie charitable donations, it is problematic to treat government support as an exogenous variable (Payne, 2009). We expect regression models and specifications that deal with omitted variable bias and endogeneity to estimate more and stronger crowding-out effects.

We test two hypotheses on regression models and model specification. First, we expect that crowding-out estimates are stronger in empirical specifications that account for time-invariant omitted variables. A simple OLS regression estimates the relation between both the level and the change in government support and private donations. Fixed-effects specifications include dummies for the units of analysis, holding all time-invariant factors constant. Most of these specifications include fixed effects for organizations (reducing bias caused by organizational size, mission, etc.), but studies with other units of analysis can include fixed effects for states or districts (reducing bias caused by population characteristics, geographical features, etc.). A first-difference estimation, regressing the changes in donations on the changes in government support, is a similar way to deal with endogeneity. Note that fixed-effects and first-difference specifications do not account for omitted variables that change over time. This can be solved by including a lagged dependent variable as a predictor in the model, but estimating both fixed effects and lagged dependents in one model comes with new (and problematic) assumptions (Angrist & Pischke, 2009: 245). Simply using a lagged government support measure as independent variable might mitigate, but not solve the bias caused by time-variant omitted variables.

Second, Payne (2009) argues that empirical specifications measuring only the exogenous part of government support, including two-staged least

squares regression (2SLS), lead to less biased estimates. Instrumental variable regression is a way to deal with the endogeneity problem, using predictor variables that correlate with the independent but not with the dependent variable or its error term (Morgan & Winship, 2007). In these models, government support is regressed on one or more instrumental variables (like region characteristics, organizational characteristics or measures of political power) to model the part of government support that is exogenous. In the second stage of the regression, private donations are regressed on the exogenous part of government support, hereby reducing the upward bias that is due to organizations receiving both high government support and high private donations.

H3: Studies using fixed-effects models and first-difference specifications are more likely to find crowding-out and find stronger crowding-out than studies using other model specifications.

H4: Studies using instrumental variable regression models are more likely to find crowding-out and find stronger crowding-out than studies using other regression models.

Government support

Our final set of hypotheses concerns the operationalization of the independent variable in primary studies. In experimental designs, researchers mostly simulate a government tax by imposing an involuntary contribution from participants. In non-experimental designs, we distinguish two dimensions that can raise differences.

First, measures of government support are either expenditures directly targeted at the need in society or subsidies to nonprofit organizations. Government support may have a direct effect on individual donations because people derive utility from the total amount that they contribute to the public good, either through taxes or through their own voluntary donations. However, it is unlikely that people change their behavior when they are not aware of (changes in) government support (Horne et al., 2005). Government support may also have an indirect effect on donations through the behavior of organizations, who play a crucial role because they collect donations and may increase their fundraising efforts when government support is lowered or vice versa. The latter effect has been labeled “fundraising crowd-out” and is a plausible explanation of the negative relation between government support and private donations (Andreoni & Payne, 2003, 2011; Hughes, Luksetich,

& Rooney, 2014). If organizational behavior explains changes in donations, studies that take subsidies to organizations as an independent variable provide more precise estimates that capture this effect and are more likely to yield crowding-out. Direct government expenditures include a wide range of government programs that benefit public goals, either through direct spending or through mediating organizations. Subsidies to nonprofit organizations are a more precise measure of public funding through nonprofit organizations, including contracts, the purchase of services, matching grants and unconditional subsidies. While some studies estimate crowding-out with aggregated measures of public and private funding in districts or sectors, studies that use organizational-level data are expected to find more and stronger crowding-out effects.

Second, both the central government and lower levels of government can provide support for nonprofit organizations. In the case of the US, federal grants are likely to not only have an effect on individual private donations but also on spending of lower levels of government, and both private donors and lower governments are responsive to one another. The term “joint crowd-out” refers to the collective effect of federal grants on both private and lower government support, while the direct effect of federal support on private donations is referred to as “simple crowd-out” (Steinberg, 1989, 1991; Lindsey & Steinberg, 1990). State and local governments tend to match federal grants, especially when those are targeted at specific needs and thus, private donors would not only substitute a decreasing federal government grant but also the decreasing local government support that sticks to federal money. Studies that only use a measure of central government spending or only a measure of spending at lower levels could overestimate the effect of government support because a part of the change in private donations is due to the change in spending by other levels of government. Studies that include a measure of total government support, or use a model that controls for other levels of government, are expected to provide weaker crowding-out estimates. Disentangling the effects of different levels of government is important because many governance networks are found on local levels and the provision of public services is increasingly decentralized (Klijn, 2008).

H5: Studies that measure subsidies to organizations are more likely to find crowding-out and find stronger crowding-out than studies that measure government expenditures.

H6: Studies that measure only central or only lower levels of government

support are more likely to find crowding-out and find stronger crowding-out than studies that measure all levels of government support.

Other moderators

In addition to the characteristics of studies that we have discussed thus far, there are many other characteristics that could affect estimates of the crowding-out effect. First, different types of private giving can be distinguished. Empirical studies that use financial archival data often use aggregated measures of private nonprofit revenue, which include donations from individuals, companies, foundations and other organizations. Much giving is religiously orientated, which is often not directly substitutable for government provision. Second, government support can be further specified. There is a variety of grants, purchases, subsidies and vouchers that may have different effects, and while the implicit assumption in many studies is that government funding is unconditional, aggregate measures of government support often include matching grants. Government grants may have differential effects when they are publicly announced or when they are part of a larger policy shift. Third, there might be differences across organizations. Due to the small number of studies and estimates per field we cannot distinguish between different parts of the nonprofit sector, nor can we align organizations on the extent to which they are subsidy-dependent.

Furthermore, it has been argued that crowding-out effects vary with the level of government support (Borgonovi, 2006; Brooks, 2000a, 2003a), the salience of the tax (Eckel et al., 2005), the number of other donors (Ribar & Wilhelm, 2002), the difference between public goods that are generally provided by public funding and public goods that are generally provided by private funding (Tinkelman, 2010), the linearity of the cost function of public good production (Tinkelman, 2010), the number of people that initially do not contribute to a public good (Chan, Godby, Mestelman, & Muller, 2002; Tinkelman, 2010) and substitution between nonprofit organizations or between sectors (Sokolowski, 2013; Tinkelman, 2010).

Due to data limitations or research design choices, not all of these conditions have been systematically tested. Table A in the Appendix shows the moderators that are often, sometimes or not often distinguished in previous empirical work, and whether or not these moderators are tested in this meta-analysis. The large number of possible moderators in the right-bottom cell shows that the crowding-out literature still has a long way to go after this meta-analysis.

DATA AND METHODS

The meta-analysis we present relies on a sample of previous studies on the crowding-out effect. To ensure comparability we limit our review to studies with the amount of donations of money as the dependent variable, either self-reported in surveys or observed in experiments or in archival (financial information) data, and the amount of government support as independent variable. Governments can enhance donations by matches or rebates (Eckel & Grossman, 2003; Pélouza & Steel, 2005), but our analysis is restricted to unconditional government grants.

A meta-analysis is a good way to examine differences between studies on the crowding-out effect. The term “meta-analysis” has been proposed by Glass (1976: 3) as referring to “the statistical analysis of a large collection of analysis results from individual studies for the purpose of integrating the findings.” Such analyses have become quite common in educational research, psychological research and especially in medical research, and are increasingly used in several other social science areas. Meta-analyses are useful in calculating an average of effect sizes that are found in a number of studies, and in examining differences among studies by running meta-regressions of study characteristics on the effect size.

Data were collected in two stages. In the first stage we used EndNote X7 to retrieve studies in the Web of Science database. We search for studies (1) with the term “crowding-out” in the title, keywords or abstract, or (2) that use a pair of possible formulations of the dependent and independent variable in title, keywords or abstract.¹ In the second stage we browsed the reference lists of the studies in the sample that we obtained from Web of Science to look for additional peer-reviewed journal articles that suited our criteria.

The sample contains studies with quantitative empirical research on the relation between government support and private charitable donations. We include studies with charitable donations by individuals or households, either observed or self-reported, as dependent variable. Donations should be charitable in the sense that the donors do not have a personal relation with the recipients, so studies on private transfers between households are not included in this meta-analysis. Studies measuring the incidence of donating are excluded, as we are interested in the amounts donated. The independent

¹ The search command used is: “(crowding-out OR crowding out OR crowd-out OR crowdout OR crowd out) OR ((donations OR giving) AND (government OR subsidies OR tax OR taxing OR taxes OR matching OR rebate OR rebates OR altruism)).” This command yielded 4,930 records on February 26, 2015.

variable of interest is the amount of government support to a goal or organization, either real or simulated. Government support should be unconditional, so matches, rebates and tax and price elasticities of private donations are excluded. Many studies use an aggregated measure of government support, which is one of the major weaknesses of a part of the crowding-out literature. Studies that take an aggregate measure of government expenditures are in our sample, even though these expenditures often include matching grants.²

Most studies contain multiple estimates of the relation between government support and private donations, using different regression models or specifications, treatment groups, or sub-samples. We code every estimate separately, so we obtain a sample of estimates that are clustered within studies. Besides a dichotomous variable on finding a negative or a positive correlation, we calculate a standardized crowding-out estimate: what is the change of private donations in the case of a \$1 increase in government contributions?³

Our search resulted in a set of 70 studies that matched the criteria, of which the main study characteristics and findings are displayed in the Appendix. Because most studies report different estimates of the association between government contributions and private donations we extracted a total of 422 findings of crowding-out or crowding-in. It is not possible to calculate a standardized effect size estimate for every finding, so the sample of standardized crowding-out effects includes 325 results from 54 studies that estimate the effect on private donations of a \$1 increase in government contributions.⁴

The sample of effect sizes contains a number of extreme values. To prevent

² We excluded 18 estimates from 6 studies that use only subsidies from the American National Endowment for the Arts (NEA) as independent variable, because those are matching grants by nature. Studies that use contributions from other private donors as independent variables, like large gifts from famous lead donors intended to increase fundraising success, are also excluded because we are theoretically interested in the effect of government policies.

³ In the case of an unstandardized regression or correlation coefficient of 0.5 and independent and dependent variables measured in absolute values, the estimate equals 0.5. When a treatment group donated \$20 on average while the government contribution was \$25, and the control group donated \$10 by a government contribution of \$5, the estimate equals $(20-10)/(25-5)=0.5$. We do not compute an estimate in the case of transformed variables like logarithmic variables (58 estimates from 11 studies) or relative measures (6 estimates from 2 studies), neither do we include an estimate if the model includes a quadratic term of government support (13 estimates from 2 studies).

⁴ Missing values on the standardized estimate are not randomly distributed in the sample. Independent samples t-tests show that studies with non-experimental data, studies from Europe, other specifications than fixed-effect or first-differences, regression models without instrumental variables and studies that use only one level of government as independent variable are less likely to report a standardized crowding-out effect size estimate.

these outliers from having a disproportionately large influence on the results the one percent lowest values are given the value of the first percentile while the one percent highest values are set on the value of the ninety-ninth percentile. This procedure is known as “Winsorizing.” As opposed to trimming, where the lowest and highest values are deleted, this method treats the data for outliers while leaving all relevant data points in the sample, making the descriptive and regression results more robust (Tukey, 1962: 17-19).

The sample includes 262 findings of a negative correlation between government support and charitable donations, and 160 findings of a positive correlation. Figure 1 graphically displays all standardized crowding-out estimates after treating the data for outliers, each horizontal line representing one study. Table 1 contains descriptive statistics for findings of crowding-out or crowding-in, the crowding-out effect estimate and the study characteristics that are used in the analyses. The median is -0.18 and the robust unweighted mean is -0.17, with a 95 percent confidence interval between -0.25 and -0.09, indicating that a \$1 increase in government support is associated with a \$0.17 decrease in private charitable donations across all studies.

We test our hypotheses in two stages. First, we examine H1 in a comparison of mean findings in experimental and non-experimental studies. An Analysis of Variance (ANOVA) is used to test whether differences between the groups are statistically significant.

Experiments differ from other studies in many ways, so in the second stage we test the remaining hypotheses for experimental and non-experimental research designs separately. We run logistic regression analyses on the binary variable of crowding-out (value 0) vs. crowding-in (1) as well as linear regression analyses on the smaller sample of standardized effect size estimates. H3 to H6 are only tested with non-experimental studies because experimental designs do not vary on these dimensions.

The probability of finding a positive association between government support and charitable giving is estimated with a logit model. Because estimates are clustered within studies we allow intercepts to vary across studies, examining the model

$$P(\text{crowding-in})_{ij} / (1 - P(\text{crowding-in})_{ij}) \\ = \beta_0 + \beta_1 X_{1ij} + \beta_2 X_{2j} + \dots + \beta_k X_{kij} + u_j + e_{ij}$$

where $P(\text{crowding-in})_{ij}$ is the probability of finding a positive correlation of

Figure 1: Dot graph of crowding-out effect estimates per study

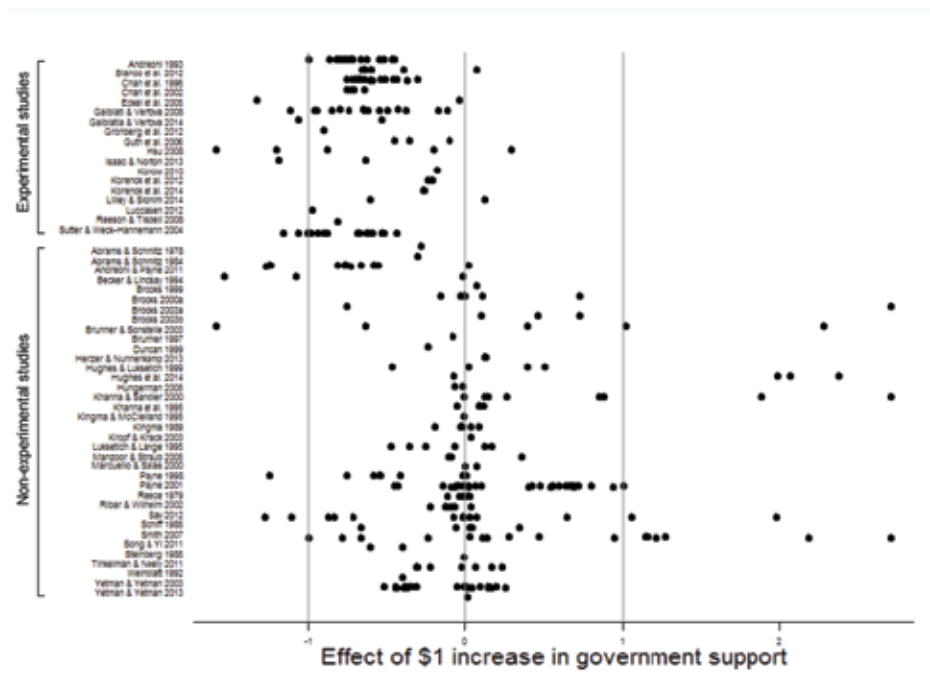


Table 1: Descriptive statistics

	N	Mean	SD	Min	Max
Full sample					
Crowding-out (0) vs. crowding-in (1)	422	0.379	0.486	0	1
Crowding-out effect estimate	325	-0.170	0.707	-1.580	2.707
Experimental study (no/yes)	422	0.268	0.443	0	1
Experimental studies					
Less generous welfare state (no/yes)	113	0.575	0.497	0	1
Year of publication	113	2003.425	7.250	1993	2014
Sample size (ln)	113	4.748	0.696	3.611	6.908
Non-experimental studies					
Less generous welfare state (no/yes)	306	0.941	0.236	0	1
Fixed-effects or first-difference (no/yes)	306	0.412	0.493	0	1
Instrumental variable (no/yes)	306	0.265	0.442	0	1
Subsidies to organizations (no/yes)	306	0.814	0.390	0	1
Only central government (no/yes)	306	0.101	0.302	0	1
Only lower government (no/yes)	306	0.052	0.223	0	1
Year of publication	306	2001.324	8.591	1978	2014
Sample size (ln)	306	5.540	2.191	0.693	14.864

the i th estimate in the j th study, β_0 the baseline intercept, β_k the regression coefficient of the k th independent variable, u_j the study-specific intercept, and e_{ij} the error term for each estimate. We report odds ratios, to be interpreted as the ratio between the odds of finding crowding-in vs. the odds of finding crowding-out. An odds ratio of 1 means that the probabilities are equal, an odds ratio below 1 means a higher probability of finding crowding-out, an odds ratio higher than 1 means that the probability of finding crowding-in is higher.

Correlates of standardized crowding-out effect estimates are estimated by linear Generalized Least Squares (GLS) regression models with the crowding-out estimate as the dependent variable and different study characteristics as the independent variables,

$$Y_{ij} = \beta_0 + \beta_1 X_{1ij} + \beta_2 X_{2j} + \dots + \beta_k X_{kij} + u_j + e_{ij}$$

where Y is the effect of a \$1 increase in government support on the amount donated.

Note that some X s only vary across studies (e.g. welfare state type) and some vary both across and within studies (e.g. the use of fixed-effects regression). Hausman tests are not statistically significant, suggesting that a random-effects specification is appropriate here.

The sample includes estimates in different parts of the voluntary sector. The sample includes 18 studies that estimate effect sizes in the field of arts and culture, 10 in the field of education, 1 study in the field of environment and animals, 7 in the health sector, 8 on international aid, 12 studies that have estimates on social services, 3 on religion, 21 studies that estimate effect sizes on an aggregated measure of giving in different sectors, and 15 studies where the receiving sector is undefined. Comparing the differences between those fields would increase our understanding of varying effects of government efforts, but the numbers of studies and estimates in each field are too small to make reliable claims.

In order to test H2 on differences between welfare state regime types we classify the United States, Canada, the United Kingdom and Australia as less generous welfare states. The only cross-country study in the sample (Sokolowski, 2013) is excluded from the regression analyses.

We include two control variables. The first control is the year of publication because the correlations of our variables of interest could be due to period effects. The second control variable is the sample size, which is often

used in meta-analyses as an indicator of the statistical power of the estimate (Borenstein, Hedges, Higgins, & Rothstein, 2009).⁵ We take the natural logarithm because the distribution of sample sizes is highly skewed.

RESULTS

Data source

Table 2 displays the means of our dependent variables for experimental and non-experimental studies. In line with H1, estimates from experiments show more and stronger crowding-out estimates. There are only 5 crowding-in estimates with experimental data in the sample, all from different studies, representing 4 percent of all experimental estimates. In non-experimental studies, there are as many crowding-out estimates as crowding-in estimates. In experiments a \$1 increase in government support is associated with a \$0.64 decrease in private donations on average (which is significantly different from zero with a 95 percent confidence interval between -0.70 and -0.58), while archival or survey data analyses find a mean increase of \$0.06 (not significantly different from zero with a 95 percent confidence interval between -0.04 and 0.15). The differences between experiments and non-experiments are statistically significant.

Table 2: Mean findings for experimental and non-experimental research designs

	Experimental studies	Non-experimental studies	Sig.
Crowding-out (0) vs. crowding-in (1)	0.044 (0.019)	0.502 (0.028)	***
Crowding-out effect estimate	-0.643 (0.031)	0.056 (0.049)	***

Differences between groups are tested with one-way ANOVA. SD values are given in parentheses.

***p < .01.

⁵ Some studies do not report sample sizes for each estimate because it uses sub-samples for different estimates. In those cases we calculated an approximate sample size based on the size of the whole sample.

Sample country

Table 3 reports the odds ratios of the logistic regression models, and Table 4 displays the regression coefficients from the GLS models. Non-experimental research designs in less generous welfare states are less likely to find positive associations between government support and private donations. The odds ratio becomes 0.60 and 0.46 when controlling for regression model and specification (Table 3, Model VI) and type of government support (Table 3, Model V), indicating that studies from less generous welfare states have a predicted probability of 19 percent to estimate crowding-out. The differences are not statistically significant due to the large standard errors. Regarding the effect size (Table 4), experimental estimates from less generous welfare states are 0.16 more strongly positive than estimates from more generous welfare states (Table 4, Model III), and non-experimental estimates from less generous welfare states are 0.19 higher (Table 4, Model IX). The GLS coefficients are in the opposite direction of what we expected.

Regression model and specification

Our hypotheses predict that crowding-out is stronger in models and specifications that account for endogeneity. Fixed-effects or first-difference specifications are between 3.5 and 4 times more likely to find a positive association between government support and charitable donations (Table 3), which is contrary to the expectation. In line with our hypothesis, instrumental variable models more often find crowding-out. In the full model (Model VIII) the odds ratio is 0.46, indicating that instrumental variable analyses have a predicted probability of 19 percent to find crowding-out, which is not statistically significant. In the linear regression (Table 4), the differences between models and specifications are small and not statistically significant. There is a large variance in instrumental variable regression estimates: the standard deviation of crowding-out effect size estimates is 0.91 for these models. It is likely that crowding-out findings strongly depend on the instruments that are used.

The unexplained between-study variance ρ does not substantially decrease in models including variables on regression model and specification. The use of fixed-effects, first-difference and instrumental variables varies both between and within studies and does not explain much of the heterogeneity in crowding-out estimates across studies.

Government support

The expectation in H5 that government subsidies to organizations have a stronger negative effect than direct expenditures must be rejected with our data. Estimates obtained for levels of subsidies as independent variables are 7.9 times more likely to find crowding-in than estimates that use direct government expenditures (Table 3, Model VII), which is contrary to the hypothesis. In the linear regression model the coefficient is positive too, although it is not significantly different from zero.

Contrary to the expectation, crowding-out estimates are not stronger when different levels of government spending are measured. Instead, estimates with a measure of only central government spending find more (Table 3, Model VIII) and stronger (Table 4, Model VIII) positive effects. H6 is rejected.

The intraclass correlation ρ does not substantially decrease when differences between measures of government support are included in the model.

Robustness check

As a robustness check we reran our analyses several times, each time excluding one study. The data are already treated for outliers (see under Data and Methods), but studies with extreme values can still have a disproportionately large influence on the results.

The mean effect size estimate of $\bar{x}=-0.17$ has a 95 percent confidence interval from -0.25 to -0.09, and excluding influential studies does not result in a mean outside this range.

The differences between the means of experimental and non-experimental designs are large and robust. Most results from the random-effects models are robust against excluding one of the studies in the sample too, with two exceptions. First, when excluding a study by Hughes et al. (2014) GLS regression coefficient of fixed-effects and first-difference models becomes more strongly negative ($\beta=-0.20$ in the full model, $p=0.09$). Hughes and colleagues find strong positive coefficients in their fixed-effects models with archival data on symphony orchestras. Second, excluding one of the studies by Brooks (2000b) makes the GLS regression coefficient of subsidies to organizations moderately negative ($\beta=-0.19$, $p=0.60$). Using longitudinal data, Brooks estimates coefficients close to zero but also one positive coefficient of 0.73 among arts and cultural organizations.

In sum, there is robust evidence that experimental designs find more and

Table 3: Logistic regression results on crowding-out (0) vs. crowding-in (1), random effects

	Experimental studies			Non-experimental studies					
	I	II	III	IV	V	VI	VII	VIII	IX
Generous welfare state		ref.	ref.		ref.	ref.	ref.	ref.	ref.
Less generous welfare state		1.229 (1.405)	0.988 (1.182)		0.935 (1.024)	0.604 (0.664)	0.464 (0.480)	0.436 (0.464)	0.444 (0.477)
Other specifications					ref.	ref.	ref.	ref.	ref.
Fixed-effects or first-difference					4.012** (2.445)	3.653** (2.112)	3.547** (2.091)	3.460* (2.197)	3.460* (2.197)
No instrumental variables					ref.	ref.	ref.	ref.	ref.
Instrumental variables					0.714 (0.358)	0.530 (0.263)	0.460 (0.234)	0.460 (0.236)	0.460 (0.236)
Direct government expenditures					ref.	ref.	ref.	ref.	ref.
Subsidies to organizations					7.940*** (5.890)	9.778*** (8.405)	9.388** (8.399)	9.388** (8.399)	9.388** (8.399)
Both levels of government					ref.	ref.	ref.	ref.	ref.
Only central government					3.540* (2.518)	3.555* (2.543)	3.540* (2.518)	3.555* (2.543)	3.555* (2.543)
Only lower government					0.953 (1.300)	0.947 (1.284)	0.953 (1.284)	0.947 (1.300)	0.947 (1.300)
Year of publication			1.293* (0.181)						
Sample size (ln)			1.455 (0.829)						
(Constant)	0.038*** (0.027)	0.034*** (0.034)	0.000* (0.000)	0.790 (0.253)	0.840 (0.879)	0.848 (0.870)	0.252 (0.270)	0.220 (0.246)	0.000 (0.000)
Between-study SD	1.026	1.026	0.001	1.786	1.785	1.738	1.534	1.559	1.571
Rho	0.242	0.242	0.000	0.492	0.492	0.479	0.417	0.425	0.429
No. of studies	20	20	20	49	49	49	49	49	49
Observations	113	113	113	306	306	306	306	306	306

Odds ratios are reported. SE values are given in parentheses; * p < .10; ** p < .05; *** p < .01

Table 4: GLS regression results on crowding-out effect size estimate, random-effects

	Experimental studies			Non-experimental studies					
	I	II	III	IV	V	VI	VII	VIII	IX
Generous welfare state		ref.	ref.		ref.	ref.	ref.	ref.	ref.
Less generous welfare state		0.053 (0.122)	0.155 (0.120)		0.171 (0.424)	0.140 (0.429)	0.170 (0.431)	0.168 (0.435)	0.193 (0.433)
Other specifications						ref.	ref.	ref.	ref.
Fixed-effects or first-difference					0.083 (0.129)	0.083 (0.129)	0.075 (0.128)	0.049 (0.134)	-0.069 (0.151)
No instrumental variables					ref.	ref.	ref.	ref.	ref.
Instrumental variables					-0.019 (0.116)	-0.019 (0.116)	-0.042 (0.119)	-0.041 (0.121)	-0.005 (0.122)
Direct government expenditures							ref.	ref.	ref.
Subsidies to organizations							0.116 (0.183)	0.116 (0.235)	0.047 (0.280)
Both levels of government								ref.	ref.
Only central government								0.352* (0.209)	0.359* (0.208)
Only lower government								-0.052 (0.343)	0.079 (0.355)
Year of publication			0.014 (0.009)						0.011 (0.010)
Sample size (ln)			-0.175** (0.081)						-0.060* (0.035)
(Constant)	-0.612*** (0.058)	-0.643*** (0.094)	-27.497 (17.635)	0.048 (0.066)	-0.119 (0.419)	-0.118 (0.419)	-0.232 (0.457)	-0.235 (0.483)	-21.081 (20.483)
Between-study SD	0.183	0.194	0.175	0.228	0.235	0.224	0.214	0.244	0.229
Rho	0.285	0.309	0.264	0.115	0.122	0.112	0.103	0.131	0.116
No. of studies	18	18	18	36	36	36	36	36	36
Observations	105	105	105	220	220	220	220	220	220

SE values are given in parentheses; * p < .10; ** p < .05; *** p < .01

stronger crowding-out, that fixed-effects of first-difference models less often estimate crowding-out, that studies using subsidies to organizations as measure of government support are more likely to find crowding-out, and that studies using a measure of central government support find more and stronger crowding-out estimates.

DISCUSSION AND CONCLUSIONS

In previous research, questions have been posed about the effectiveness of new forms of governance with larger roles for nonprofit organizations in the creation and implementation of public services (Ansell & Gash, 2008; Milward & Provan, 2003; Smith & Lipsky, 1993). In order to understand the contextual dynamics of effective governance, there is a need for robust evidence on the effects of changing government spending on fundraising income. Despite a large number of empirical studies there is no decisive evidence for government support to crowd out private charitable contributions. About two-thirds of the findings in our meta-analysis show a negative correlation between government support and charitable donations, while one third finds a positive correlation.

Payne (2009) argues that research on the relation between government support and charitable donations suffers from endogeneity. One way to establish causality is through experimental research designs, and our analysis shows that these designs find more and stronger crowding-out effects than studies using archival or survey data. While experiments show that each dollar of government support crowds out \$0.64 of private donations, a dollar increase in government support in non-experimental data from surveys, financial information forms or other archival data is associated with a slight *increase* in voluntary contributions on average. Our analysis shows that there is incomplete crowding-out and that the pure altruism model, in which each dollar of mandatory contributions leads to a dollar reduction in voluntary contributions, should be reconsidered. The pure altruism model makes a number of assumptions about the situation in which the government and private donors contribute to a public good, and crowding-out findings depend on the extent to which empirical studies relax these assumptions (Tinkelman, 2010). In experiments people have full information on the level of government contributions, decide on money that is not their own, are sensitive to social cues because they are aware of taking part in a study,

and are often undergraduate students that differ in their prosocial behavior and reactions to experimental manipulation (Henrich et al., 2010). There is an ongoing debate about the extent to which findings from laboratory experiments can be generalized to natural settings (Camerer, 2015; Galizzi & Navarro-Martinez, 2015; Levitt & List, 2007) and the large difference in our meta-analysis sample between the estimates obtained in experiments and other types of data emphasizes the importance of this debate.

One could argue that experimental designs provide cleaner estimates of the causal relation because they rule out the interference of other variables. If endogeneity explains why experimental findings differ from other findings, we would observe that regression models and specifications that effectively deal with this issue produce stronger crowding-out estimates than other regression models. Our results do not confirm this line of reasoning. Neither fixed-effects or first-difference specifications nor the use of instrumental variables are robustly linked with stronger crowding-out. It is likely that findings in instrumental variable models are highly dependent on the measures that are used as instruments. Similar measures of organizational output and region characteristics are used by some studies as instruments for government support (Brooks, 1999; Khanna & Sandler, 2000; Payne, 2001) and by another study as instruments for private giving (Becker & Lindsay, 1994). Hughes and Luksetich (1999) use the same set of variables as instruments for both public and private funding sources in different 2SLS regression models. If a prerequisite for a valid instrumental variable is that it is correlated with X but not with Y or its error term (Morgan & Winship, 2007), it is striking that the same kind of variables are used for both government support and charitable giving. Researchers should be very careful in applying these techniques, and preferably use a range of different models, specifications and instrumental variables to estimate the effect of government support in a certain dataset.

Our results also challenge the argument of indirect crowding-out, which means that the fundraising behavior of organizations partly explain why people change their donations after government investments or budget cuts (Andreoni & Payne, 2003, 2011; Hughes et al., 2014). Subsidies to organizations are much more likely to crowd in donations than direct government expenditures, but they do not lead to *stronger* crowding-in effects on average. A possible explanation for this result is that the effect is non-linear, with smaller subsidies enhancing donations and larger subsidies discouraging them (Borgonovi, 2006; Brooks, 2000a, 2003a). This also means that subsidizing

does not make organizations dependent on public funding, but rather seems to encourage revenue diversification at the organizational level. Considering previous arguments that governance networks benefit from resource-rich environments (Milward & Provan, 2000) and that organizations with a diversified revenue mix tend to be financially stable (Carroll & Stater, 2009), subsidizing the nonprofit sector could strongly improve the chances of fruitful public-private partnerships.

Our analyses show that measures of central government support are positively related to charitable donations, while measures of multiple governmental levels are negatively related to giving. This is contrary to what we expect from models developed by Steinberg (1989, 1991) and Lindsey and Steinberg (1990). If there would be a “flypaper effect,” meaning that federal funding induces support from lower levels of government to the same public good, studies of central government support would underestimate the total crowding-out effect. The results from this meta-analysis contradict this argument. Federal policy programs turn out to be effective in stimulating private giving, while policy programs on local levels, which often involve nonprofit actors, are difficult to fund through a mix of public and private funding. This raises the question how effective local policy makers are in establishing fruitful collaborations with nonprofit actors, which is an emerging topic in an era in which public services are increasingly decentralized (Klijn, 2008).

Our analysis suffers from a few limitations. First, there are more differences between research designs than we accounted for in this paper. A common critique on meta-analyses is that they compare apples and oranges by including findings that diverge in many more ways than can be tested for (Borenstein et al., 2009: 379-380; Petticrew & Roberts, 2006: 203-204; Wolf, 1986: 14-15). Without doubt, different measures of the dependent and independent variable lead to different findings. In future research, comparing the effects of different types of government support and on different types of organizations would add much to our understanding about nonprofit financing across society. The most important difference we found is the one between experimental and non-experimental studies, and there are numerous differences between these two approaches that cannot all be examined by meta-analytical techniques. In the current analysis we cannot establish with certainty to what extent the stronger crowding-out results in experiments are due to the information that is provided, the endowment participants receive, demand effects, subject pool composition effects or reduced endogeneity. Systematic comparisons between data that vary on dimensions

that are not often tested in previous empirical research (see the Appendix, Table A) could provide more insight. Furthermore, the current analysis concerns the relationship between funding sources without paying attention to the actors involved or the governance processes behind it. Although the environmental dynamics are important for organizations and the analysis is valuable as such, reactions on funding streams may depend on many other factors like institutional characteristics, management styles and relations between the actors involved (Ansell & Gash, 2008; Milward & Provan, 2003). More research would be necessary to shed light on other factors that moderate crowding-out effects.

A second important limitation is that the estimates in our meta-analysis are not necessarily a random sample. Weak or non-significant results are generally more likely to remain unpublished (Borenstein et al., 2009: 277-292; Francis, 2012; Petticrew & Roberts, 2006: 230-235; Rosenthal, 1984: 125; Stanley, 2005) and our search technique excludes findings from books and “grey literature.” Although the findings presented here are robust, our analyses concern a possibly biased sample of all crowding-out estimates that empirical research is able to measure. Being a generally recognized problem of scientific publishing, publication bias is less likely to be a problem in crowding-out research because null findings in this area have important policy implications. An analysis of unpublished studies could be added in order to examine this bias, which is beyond the scope of the current article.

Despite these limitations, this paper makes an important contribution to the literature on the interaction between organizations and their environment. In field research situations, where different environmental processes are at play, individual giving is generally not strongly affected by varying levels of government support. However, private donors are responsive to changing government support under certain circumstances. When people are aware of government budgets they might change their donations, so the effects of public policy largely depend on information flows. In general we advise policy makers to carefully consider the societal context before deciding to reduce public spending, since budget cuts mostly decrease total funding for public goods. This has important consequences for governance styles in which the government collaborates with nonprofit actors, like nonprofit contracting (Smith & Lipsky, 1993), “collaborative governance” (Ansell & Gash, 2008) and interorganizational networks (Milward & Provan, 2003).

There is a widespread belief among politicians and intellectuals that government expenditures suppress private participation, an assumption that

lies behind policy decisions in which the government cuts its spending and aims to shift public services towards nonprofit organizations that are largely dependent on private funding. The current meta-analysis shows that in most situations, private charitable donations are not likely to be crowded out by government support and that each dollar of extra public funding increases total contributions to the public good. When governments are able to maintain high levels of public funding, they may continue to seek collaborations with nonprofit actors as complementary in the funding and implementation of public services. Instead of substituting each other, there is ample opportunity for government and nonprofits to jointly enhance the scope and quality of public services in different organizational arrangements.