NBER WORKING PAPER SERIES

CROWDING IN WITH IMPURE ALTRUISM:
THEORY AND EVIDENCE FROM VOLUNTEERISM IN NATIONAL PARKS

Matthew Kotchen
Katherine R.H. Wagner

Working Paper 26445
http://www.nber.org/papers/w26445

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
November 2019

We are grateful to Ted Bergstrom, Nathan Chan, Joshua Gottlieb, Catie Hausman, Grant Jacobsen, Erin Mansur, Todd Sandler, Joe Shapiro, Lise Vesterlund, and Casey Wichman for helpful comments and discussions, along with seminar participants who commented on earlier versions of this research at Carnegie Mellon University, Columbia University, McGill University, London School of Economics, UC Berkeley, UC Santa Barbara, University of Connecticut, University of Pittsburgh, Yale University, and the 2019 NBER Summer Institute. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Crowding In with Impure Altruism: Theory and Evidence from Volunteerism in National Parks
Matthew Kotchen and Katherine R.H. Wagner
NBER Working Paper No. 26445
November 2019
JEL No. H40,H50,Q26

ABSTRACT

This paper makes three contributions to the literature on private provision of public goods. First, we identify limitations of the frequently used specification test that distinguishes between the standard models of pure and impure altruism based on the extent of crowding out. While the literature takes as given the result that crowding out should be less with impure altruism compared with pure altruism, we show that, in general, it can be either more or less. Second, we propose a more general test based on the presence of crowding in, rather than the extent of crowding out. Third, we provide empirical evidence. Using a unique panel data set on volunteerism in U.S. National Parks, we estimate the causal effect of changes in public funding within parks on the amount of within-park volunteerism. The overall finding is that each additional dollar of public expenditure crowds in 27 cents worth of volunteerism on average. We show how the estimates of crowding in, along with heterogeneity based on park and volunteer hour types, are theoretically consistent with the mainstay model of impure altruism.

Matthew Kotchen
School of Forestry & Environmental Studies,
School of Management,
and Department of Economics
Yale University
195 Prospect Street
New Haven, CT 06511
and NBER
matthew.kotchen@yale.edu

Katherine R.H. Wagner
Department of Economics
Yale University
28 Hillhouse Ave
New Haven, CT 06511
katherine.wagner@yale.edu

1 Introduction

How does government provision of a public good affect private provision? The question is central to public economics. Implementing the optimal supply of public goods through government policy requires an understanding of the behavioral responses to public provision. If, for example, public supply crowds out private provision, then failing to account for the behavioral response means that the intended level of the public good will come up short. If, however, there is crowding in, then public provision can be used to leverage greater private contributions. Theoretical and empirical studies of crowding effects also provide the basis for an extensive literature that seeks to identify the underlying motives of charitable behavior. In particular, the extent of crowding out is widely used as the standard test between the models of pure and impure altruism, which provide different explanations for why individuals engage in charitable activity.

In this paper, we identify limitations of the theoretical foundation for using crowding out as a specification test between pure and impure altruism. The literature takes as given the result that crowding out should be less with impure altruism compared with pure altruism, but we show that it can be more or less. Indeed, we show that the standard specification test is based on an implicitly assumed special case. We propose an alternative and more general test based on the presence of crowding in, rather than the extent of crowding out. We then illustrate the test in an empirical setting that seeks to overcome many of the challenges in previous studies that estimate crowding effects. We take advantage of a unique panel data set on volunteerism in U.S. National Parks to estimate the causal effect of changes in public funding to a park on the amount of within-park volunteerism, along with heterogeneous effects by park and volunteer-hour type.

The seminal theory on private provision of a public good establishes the basic framework for crowding out (e.g., Warr 1982; Roberts 1984; Cornes and Sandler 1985; Bergstrom, Blume and Varian 1986). With a pure public good, an individual's own provision and public provision are perfect substitutes, and the private and public goods over which individuals obtain utility are both assumed to be normal (i.e., not inferior). One consequence is that public good provision financed through lump-sum taxation crowds out private provision one-for-one, assuming an interior solution. Despite the influence and theoretical appeal of these models, many of the results, including the notion of complete crowding out, are viewed as limited because of their empirical implausibility. Indeed, Andreoni (2006) summarizes many of the implications associated with complete crowding out as a classic *reducto ad absurdum*.[1]

[1] Specifically, Andreoni (2006) argues that, "If we are going to accept complete crowding out, we also need to believe in near complete crowding of any government gifts to charity, that only the very richest are giving, that redistributions of income are neutral as long as people are giving to charities, and that even 'distortionary' taxes may be non-distortionary" (p.1219).

Other models of privately provided public goods seek to reconcile the apparent disconnect between theoretical predictions and empirical observations. In many cases, a broader formulation of individual preferences is used to help explain why crowding out might be less than complete. One of the most influential is Andreoni's (1989, 1990) model of impure altruism. In addition to enjoying the benefits of a public good, individuals are assumed to enjoy a private "warm-glow" benefit from the act of their own giving. This setup implies that private and public provision are no longer perfect substitutes, and a foundational result for much of the literature is that crowding out becomes incomplete (i.e., less than one-for-one). A more general formulation is also found in Cornes and Sandler's (1984, 1994) model of an impure public good, which is based on joint production of private and public characteristics.[2]

There exists a substantial empirical literature on crowding out that tests between the models of impure altruism and pure altruism (i.e., the pure public good model). As mentioned above, underlying much of this research is the theoretical prediction that crowding out should be less with impure altruism than with pure altruism. In an early and influential contribution, Kingma (1989) estimates incomplete crowding out of private contributions to public radio stations, and because of this, the results are generally interpreted as consistent with the model of impure altruism. Building on the same theoretical and empirical approach, other papers estimate crowding out in a variety of contexts (e.g., Ribar and Wilhelm 2002; Okten and Weisbrod 2011; Borgonovi 2006; Hungerman 2005; Payne 2009; Andreoni and Payne 2003, 2011). Another strand of the literature uses controlled laboratory experiments to estimate crowding out and test between the models of pure and impure altruism (e.g., Andreoni 1993; Eckel, Grossman and Johnston 2005; Ottoni-Wilhelm, Vesterlund and Xie 2017).

The majority of studies find evidence of crowding out, albeit less than complete. A recent meta-analysis finds that crowding out is more likely to arise in experimental studies, whereas observational studies are more clustered around the finding of no effect (de Wit and Bekkers 2017). Figure 1 shows the probability density functions for the estimated crowding effects separately for experimental and non-experimental studies. In total, 38 percent of the estimates have a finding of crowding in rather than crowding out. In these cases, because the current interpretation of the standard model of impure altruism does not admit crowding in, the results are typically described as outliers or interpreted in the context of models that account for additional elements (e.g., Khanna and Sandler 2000; Payne 2001; Andreoni and Payne 2011). With respect to individual contributions, these explanations include the possibility that public provision signals quality in a way that makes private contributions more attractive to donors, or perhaps increases the scale or

[2]Joint production is also consistent with other approaches for modeling charitable behavior based on reputation (Hollander 1990; Harbaugh 1998), signaling about income (Glazer and Konrad 1986), and environmentally friendly consumption (Kotchen 2005, 2006).

scope of fundraising.[3]

In the first part of this paper, we reconsider the theoretical basis for crowding out in the standard model of impure altruism, and we come to different conclusions. We show that the standard normality assumptions with impure altruism are not directly comparable to those for the pure public good model, because the constrained setup with impure altruism creates a divergence between normality based on primitive preferences and the income effect on demand functions.[4] One consequence is that assuming normality in the usual sense of preferences means that crowding in, along with crowding out, is a plausible consequence with impure altruism. Nevertheless, this possibility appears to be either overlooked or implicitly assumed away, with important implications for interpretation of the model.

More important, we show that, in general, crowding out can be either more or less with impure altruism compared to pure altruism, even under impure altruism's standard set of assumptions. The relative magnitudes depends in large part on the degree of Hicksian complementarity or substitutability between the jointly produced private and public characteristics (i.e., the public good, and warm glow or some other private benefit). This means that the standard specification test in the literature for distinguishing between pure and impure altruism does not hold in general, and we show how previous analyses are based on an implicitly assumed special case of additive separability in the utility function. Finally, we propose an alternative and more general specification test based on the presence of crowding in, which we show is consistent with impure altruism, but not pure altruism.

In the second part of the paper, we estimate the crowding effect in an empirical setting that allows us to illustrate the new specification test. To do so, we exploit administrative data from the Volunteers-In-Parks (VIP) program of the U.S. National Park Service (NPS) from the years 1998 to 2013. The data set includes the total number of volunteer hours in all NPS units, along with the type of activity associated with each hour (e.g., resource conservation, indoor administration, etc.). We combine these 16 years of volunteer data with other sources of data on annual park budgets and visitation. Our primary objective is to estimate the causal effect of changes in federal funding within parks on within-park volunteerism. A unique feature of the focus on volunteerism in national parks, rather than monetary contributions, is that private provision of the public good is clearly associated with a jointly produced private benefit. Volunteerism contributes to the public good of conservation, while simultaneously creating opportunities for *in situ* park enjoyment. Un-

[3]We focus here on the private provision of individuals. Another strand of the public finance literature studies crowding in or out at different levels of government. In this context, crowding in is sometimes referred to as the "flypaper effect" in recognition that public money tends to stick where it hits (Hines and Thaler 1995).

[4]This results emerges from a comparison of the analytical approaches of Andreoni (1989, 1990) and Cornes and Sandler (1984, 1994).

like most other studies of crowding effects, our study therefore takes place in a setting where impure altruism is the expected model *ex ante*, rather than the model deduced from the estimates *ex post*.[5] Hence a finding of crowding in is consistent with our new specification test for distinguishing between the models of pure and impure altruism.

The empirical setting also has a number of unique advantages that help to overcome challenges that arise in existing studies of crowding effects. First, all of the park units in our sample are managed in the same way through the NPS, and this adds a unique degree of homogeneity across the units of analysis. Most other studies rely on cross-sections or panels of charities that provide different public goods and operate in different ways.[6] Second, federal funding accounts for the vast majority of the operating budget within parks, thereby reducing concerns about endogenous fundraising from outside sources and funding from other levels of government. Third, this is the first study we are aware of to examine crowding out or in of time rather than money. Fourth, because individuals generally volunteer in only one park, our analysis is not subject to concerns about substitution between units of study masked by aggregate data. Fifth, the breadth and length of our panel means that we can estimate within-unit crowding effects over a longer period of time than most other studies. Finally, the detailed data on volunteer hours broken down by type of activity enables a unique opportunity to examine heterogeneous effects informed by theory.

Our preferred estimates are based on a shift-share instrumental variables approach. Specifically, we instrument for changes in park funding with a Bartik-style interaction between individual park budgets three years prior to our period of analysis and changes in Congressional support for conservation issues, as measured by the League of Conservation Voters. We find that, on average, a \$1,000 increase in a park's annual funding increases volunteerism by 12.4 hours. Using a standard wage rate for conversion, this translates into crowding in of 27 cents for every additional dollar of public expenditure. We show how the overall results are theoretically consistent with the model of impure altruism, as are the heterogeneous results that we find based on volunteer-hour and park types. In particular, we find evidence of crowding in that is greater for volunteer hours and parks that are more conservation and outdoors oriented—that is, in circumstances where the joint production of public and private benefits are most likely to arise.

The remainder of the paper proceeds as follows. The next three sections develop our theoretical analysis. We begin with background on the standard model for impure altruism. Sections 3 and 4 revisit the approach for analyzing crowding out, make direct com-

[5] Although the jointly produced private benefit of impure altruism is usually characterized as a warm glow, the model is equivalent to that of an impure public good (Cornes and Sandler 1984, 1994) that can apply to any private benefit, such as *in situ* park enjoyment in this case.

[6] While most studies in the literature on crowding effects focus on social services (Andreoni and Payne 2013), one distinguishing feature herein is a focus on behavior related to environmental and natural resource conservation.

parisons to show why the new set of results differ, and put forth the new specification test based on crowding in. Section 5 describes our empirical setting and data collection. Section 6 outlines our empirical strategy. Section 7 reports the estimation results, considers robustness checks, and discusses possible alternative explanations. Section 8 provides a concluding discussion.

2 Impure Altruism: Preliminaries

This section begins with the setup and assumptions of Andreoni's (1989, 1990) seminal model of impure altruism. We then reproduce the model's well-established results on crowding out to provide points of reference for our alternative derivation and reconciliation of differing results in Sections 3 and 4.

2.1 Setup and Assumptions

Individual preferences are given by a strictly increasing and strictly quasiconcave utility function of the form $U_i(x_i, G, g_i)$, where x_i is a private good, g_i is the individual's own contribution to the public good $G = g_i + G_{-i}$, and G_{-i} is the exogenously given level of the public good provided by all others. Letting w_i denote the individual's wealth endowment, and normalizing prices to unity, the individual's utility maximization problem can be written as

$$\max_{x_i, g_i} U_i(x_i, G, g_i) \text{ s.t. } x_i + g_i = w_i \text{ and } G = g_i + G_{-i}. \tag{1}$$

The inclusion of g_i as a separate argument in the utility function is the distinguishing feature of this setup, compared with the pure public good model, where preferences are given by $U_i(x_i, G)$. Utility from g_i on its own is often interpreted as a "warm glow" or "joy of giving" that one receives from voluntarily contributing to the public good. More generally, it can represent any private benefit that arises through joint production of an impure public good (Cornes and Sandler 1984, 1994). In our empirical context, which we motivate in more detail below, the choice of g_i represents time spent volunteering in a National Park, where volunteerism provides an impure public good. The joint production in this case comes from provision of a public good (e.g., conservation), along with a private benefit of park enjoyment while volunteering *in situ*, a warm-glow, or both.

Substituting the constraints in (1) into the utility function, the individual's problem can be rewritten with a choice over the aggregate level of the public good:

$$\max_{G \geq G_{-i}} U_i(w_i + G_{-i} - G, G, G - G_{-i}). \tag{2}$$

The standard approach in the literature is to then express the solution as a function of the exogenous components of the maximand such that

$$G^* = f_i(w_i + G_{-i}, G_{-i}). \quad (3)$$

One reason for writing the solution in this way is to make direct comparisons with the pure public good model (Bergstrom, Blume and Varian 1986), in which case demand for G is written more simply as $h_i(w_i + G_{-i})$, which includes only the first argument of f_i in (3).[7] Key results for the pure public good model then follow by assuming normality of both the private and public goods, which is equivalent to assuming $0 < h_i'(w_i + G_{-i}) < 1$ for interior solutions, which we assume here and throughout. These results include a unique Nash equilibrium, neutrality of wealth redistributions, and crowding out.[8]

With the impure public good model, however, there is no direct relationship between standard notions about normality of goods based on preferences and the partial effects of the different arguments in the demand function (3). This is a subject to which we will return in greater detail below, but for the moment, we reproduce the original arguments that motivate comparable assumptions and pervade the literature on impure altruism.[9] The partial effects of f_i are typically intuited as follows. Seeking to mirror the standard normality assumption for a pure public good, the claim is that normality of x_i and G implies $0 < f_{i1} < 1$. The intuition is that an increase in w_i must be spent on both the private and public goods. A further claim is that normality of x_i and g_i implies $f_{i2} > 0$. Here the intuition follows from the thought experiment of simultaneously lowering G_{-i} and increasing w_i by the same amount. The argument is that some of the increase in w_i must be spent on both x_i and g_i, so demand for G must fall and hence $f_{i2} > 0$. Finally, it is also standard to assume that $0 < f_{i1} + f_{i2} < 1$, and this condition ensures uniqueness of the Nash equilibrium.[10]

2.2 Crowding Out and Specification Tests

The preceding set of assumptions on the partial effects of the demand function has direct implications for the model's predictions about crowding out, along with comparisons to those for pure altruism (i.e., the pure public good model). By definition, an individual's

[7] The single argument of this function, $w_i + G_{-i}$, is sometimes referred to as "full income" or "social income" because it represents one's own wealth plus the value of public good spill-ins provided by others (Becker 1974). A key feature for many of the pure public good model's results is that for interior solutions, $\partial h_i/\partial w_i = \partial h_i/\partial G_{-i}$.

[8] Neutrality refers to Warr's (1983) result showing that small redistributions of the endowments among contributors to the public good will have no effect on the total level of equilibrium provision.

[9] Andreoni (1989, 1990) provides the original analysis, and more recent applications that work through the same set of assumptions include Ribar and Wilhelm (2002) and Ottoni-Wilhelm, Vesterlund and Xie (2017).

[10] See Kotchen (2007) for a proof of equilibrium existence and uniqueness. Note that the conditions of $f_{i1} < 1$ and $f_{i1} + f_{i2} < 1$ are written with a strict inequality because of the simplifying assumption of an interior solution. Without this assumption, the inequalities would need to hold only weakly.

own contribution will satisfy

$$g_i^* = f_i(w_i + G_{-i}, G_{-i}) - G_{-i}, \tag{4}$$

and there are two notions of crowding out to consider based on this expression. First is the unfunded effect of an exogenous change in G_{-i}:

$$\frac{\partial g_i^*}{\partial G_{-i}} = f_{i1} + f_{i2} - 1. \tag{5}$$

Second is a funded effect, where lump-sum taxation is used to fund the change in exogenous provision. In this case, we have

$$\left.\frac{\partial g_i^*}{\partial G_{-i}}\right|_{dG_{-i}=-dw_i} = f_{i2} - 1. \tag{6}$$

The assumption above that $0 < f_{i1} + f_{i2} < 1$ means that both (5) and (6) are between -1 and 0. This implies crowding out that is less than one-for-one in either the funded or unfunded case, and the assumption rules out the possibility for crowding in. More important, because $f_{i2} > 0$, a further prediction is that crowding out will always be less with impure altruism than with pure altruism (because in the latter case $f_{i2} = 0$). In particular, the standard theoretical result in the literature is that funded crowding out that is less than one-for-one is consistent with impure altruism but not pure altruism.

The preceding observations underlie an extensive empirical literature that tests between the motives of pure and impure altruism for private provision of public goods. Nevertheless, we aim to show in what follows that the preceding specification test for impure altruism is not a general result. In Section 3, we show that normality assumptions in the impure public good model differ from those in the pure public good model because the former's constrained setup creates a difference between normality based on primitive preferences and the income effect on demand functions. This means that assuming normality based on preferences readily admits crowding in, along with crowding out, as a plausible consequence. In Section 4, we show that the key condition underlying the specification tests—that is, $f_{i2} > 0$—does not hold in general, and follows based on the implicit assumption of additive separability of the utility function. More generally, we show that crowding out with impure altruism can be greater than or less than that with pure altruism. This renders the results of the current specification test based on the extent of crowding out inconclusive. As an alternative specification test, we propose focusing on the presence of crowding in, rather than the extent of crowding out, to help distinguish between underlying motives consistent with pure or impure altruism.

3 The Crowding Effects Revisited

In the model of impure altruism, the way that g_i enters the utility function in two different arguments implies joint production, whereby a unit of g_i jointly produces a unit of the public good and an associated private good (e.g., warm glow). Each individual's budget frontier is therefore a ray in three-dimensional utility space, as shown by segment AB in Figure 2. This feature of the model means that comparative static results, including those related to crowding effects, are more constrained than previous analyses of the impure public model indicate. In what follows, we employ an alternative approach for examining crowding effects that yields results in terms of familiar price and income responses. In particular, we use the method of comparative static analysis developed by Cornes and Sandler (1994, 1996) for the more general model of an impure public good.

The starting point is to recognize that the solution to (2) can be written alternatively and simply as a function of the exogenous parameters such that $G^* = G^*(w_i, G_{-i})$. This identifies a point on the interior of segment AB in Figure 2 that is tangent to the individual's strictly convex indifference surface.[11] Now, for the moment, consider an alternative utility maximization problem where the individual has unconstrained choices among (x_i, G, g_i) and faces a standard budget constraint with "virtual" prices and income, which can initially take on arbitrary values. The thought experiment is to decouple the linear joint production between G and g_i, such that the individual faces a standard, three-good consumer problem of the form

$$\max_{x_i, G, g_i} U_i(x_i, G, g_i) \text{ s.t. } x_i + \pi_G G + \pi_g g_i = m_i, \tag{7}$$

where the price of x_i is normalized to unity, and π_g, π_G, and m_i are standard prices for each of the goods and income, respectively. As noted, an key component of (7) is that the constraint $G = G_{-i} + g_i$ no longer applies. It follows that the unrestricted demand for each of the goods can be written as a function of the virtual prices and income. For example, demand for the public good can be written as $G(\pi_G, \pi_g, m_i)$.[12]

The bridge between utility maximization problems (1) and (7) is to set the virtual prices

[11] Comparative statics can, of course, be derived for $G^*(w_i, G_{-i})$ directly by differentiating the first-order condition to (2). This yields expressions that depend on the second- and cross-partial derivatives of the utility function, as shown, for example, in the proofs by Ottoni-Wilhelm, Vesterlund and Xie (2017). While the approach is useful for illustrating some of the different possibilities, it is limited with respect to interpretation based on the cross-partial derivatives. Indeed, standard consumer theory does not rely on super(sub)modular definitions of complements (substitutes) because these are not necessarily preserved under monotonic transformations of ordinal preferences. The Cobb-Douglas utility function used in previous studies provides an example. The approach we employ here translates the constrained comparative static results for the model of impure altruism in (1) into the familiar price and income effects of standard consumer theory.

[12] We focus on demand for G rather than g_i in order to facilitate comparisons with the standard results discussed in the previous section. One could similarly focus on $x_i(\pi_G, \pi_g, m_i)$ and $g_i(\pi_G, \pi_g, m_i)$, where the latter could be used to derive the same results that follow.

and income to define a plane in three-dimensional utility space that includes segment AB in Figure 2 and is tangent to the individual's indifference surface at her chosen allocation from (1). The plane CDE provides an illustration. Then, it follows by definition that

$$G^*(w_i, G_{-i}) = G(\pi_g, \pi_G, m_i),$$

where the virtual magnitudes themselves are a function of preferences and the exogenous parameters w_i and G_{-i}. Note that the asterisk on G^* indicates the choice constrained to the budget ray AB, while G without the asterisk indicates the unrestricted demand function. The general approach is then to recognize that

$$\frac{\partial G^*}{\partial \theta} = \frac{\partial G}{\partial \pi_g} \cdot \frac{\partial \pi_g}{\partial \theta} + \frac{\partial G}{\partial \pi_G} \cdot \frac{\partial \pi_G}{\partial \theta} + \frac{\partial G}{\partial m_i} \cdot \frac{\partial m_i}{\partial \theta}, \tag{8}$$

where θ represents one of the exogenous parameters. The key insight of (8) is that comparative statics results for the constrained problem can be interpreted in terms of familiar price and income effects in the unconstrained problem. The next step is to recognize that we can solve for changes in the virtual magnitudes using Cramer's Rule with three conditions that link virtual magnitudes to exogenous parameters. We provide these details in the Appendix.

Turning immediately to results, we begin with the effect of a change in the individual's wealth:

$$\frac{\partial G^*}{\partial w_i} = \frac{\partial g_i^*}{\partial w_i} = [(\bar{g}_{\pi_G} - \bar{g}_{\pi_g})G_{m_i} + (\bar{G}_{\pi_g} - \bar{G}_{\pi_G})g_{m_i}]\frac{1}{\Omega}. \tag{9}$$

The first equality follows from (4). The terms G_{m_i} and g_{m_i}, where subscripts denote partial derivatives, represent the income effect on unrestricted demand for the public good and the jointly produced private benefit, respectively. Assuming normality of both unrestricted goods means that both terms are positive. The terms $\bar{g}_{\pi_g}$ and $\bar{G}_{\pi_G}$, which arise after substituting the Slutsky decomposition into (8), are the compensated own-price responses, both of which are negative. The denominator $\Omega = \bar{g}_{\pi_G} + \bar{G}_{\pi_g} - \bar{g}_{\pi_g} - \bar{G}_{\pi_G} > 0$.[13] Finally, $\bar{g}_{\pi_G} = \bar{G}_{\pi_g}$ is the compensated cross-price response, which is positive or negative depending on whether the two goods are net (Hicksian) substitutes or complements, respectively.

Equation (9) begins to show how the impure altruism model can generate surprising comparative static results. It shows that assuming normality of both goods in the typical, unrestricted sense does not imply that $\partial G^* / \partial w_i$ is greater than zero. Whether G and g_i are substitutes or complements plays a role in determining the sign of (9). If they are substitutes, the expression is unambiguously positive. If they are complements, the expression

[13] Negative semi-definiteness of the matrix of compensated price responses implies that $\Omega \geq 0$, which we assume holds strictly.

can take either sign.[14]

The different possibilities reveal two notions of normality that should be distinguished when analyzing impure altruism. Normality of goods based on primitive preferences is an assumption about the locus of points consistent with a constant marginal rate of substitution. In a standard (unrestricted) utility maximization problem, this has clear implications for the income effects on demand (i.e., G_{m_i}, $g_{m_i} > 0$). With joint production, however, the relationship between normality with respect to preferences is not the same as normality with respect to income effects on demand, as shown in (9). The difference occurs in this case because the individual's choice is restricted to the budget ray in Figure 2 rather than any point on the corresponding plane. While the distinction does not arise in the pure public good model, because it is unrestricted, the implication for impure altruism is consequential: assuming normality in a comparable, unrestricted sense, does not imply that an increase in w_i will increase demand for one's own private provision.

We now turn directly to the crowding effects, and we again provide the details in the Appendix. Solving for the unfunded crowding effect yields

$$\frac{\partial g_i^*}{\partial G_{-i}} = \frac{\partial G^*}{\partial G_{-i}} - 1 = (\bar{g}_{\pi_g} - \bar{g}_{\pi_G})\frac{1}{\Omega} + \pi_G \frac{\partial g_i^*}{\partial w_i}, \tag{10}$$

and the funded crowding effect is

$$\left.\frac{\partial g_i^*}{\partial G_{-i}}\right|_{dG_{-i}=-dw_i} = \frac{\partial g_i^*}{\partial G_{-i}} - \frac{\partial g_i^*}{\partial w_i} = (\bar{g}_{\pi_g} - \bar{g}_{\pi_G})\frac{1}{\Omega} - \pi_g \frac{\partial g_i^*}{\partial w_i}. \tag{11}$$

The first equality in (10) follows from (4). The second equality in (11) uses the condition that $\pi_G + \pi_g = 1$, which follows because the virtual prices of the jointly produced products must sum to the price of joint product itself (see the Appendix). A general observation about (10) and (11) is that both expressions have indeterminate signs under either of the two normality assumptions described above. Importantly, this means that crowding out or in is admissible if we assume that $G_{m_i}, g_{m_i} > 0$, or even more restrictively that $\partial G^*/\partial w_i = \partial g_i^*/\partial w_i > 0$. In fact, the only case where crowding out is assured occurs with the funded effect and net complementarity between G and g_i.[15] The first two columns of Table 1 summarize these results and others based on (9), (10) and (11).[16]

[14] It is straightforward to show that with complements, $\Omega > 0$ requires that one of the two terms in parentheses in (9) must be positive and the other negative. The sign therefore depends on the respective differences of these terms weighted by the corresponding income effects.

[15] In a study of military expenditures among NATO allies, Murdoch and Sandler (1984) point to the potential importance of complementarity relationships for the study of crowding effects. In a companion paper, Murdoch and Sandler (1990) use the non-neutrality with joint production as part of a test between Nash-Cournot and Lindahl behavior among NATO allies.

[16] As we discuss in the next section, the results included in the table related to $\partial G^*/\partial G_{-i}$ follow immediately from adding one to both sides of (10) and rearranging terms.

4 Reconciliation and Specification Test

The previous section derives results that differ from those in the standard analysis of impure altruism. In this section, we directly compare the two approaches and establish two new results. The first is that contrary to conventional wisdom in the literature, crowding out need not be less with impure altruism than with pure altruism. The second is that the presence of crowding in (rather than crowding out) provides a more general specification test between the two models.

We first consider the effect of a change in an individual's exogenous wealth. Referring back to the notation in Sections 2 and 3, it holds by definition that $f_{i1} = \partial G^*/\partial w_i$. The standard assumption that $0 < f_{i1} < 1$ is described as a consequence of normality of the private and public goods. Our analysis shows how this is equivalent to assuming bounds on equation (9). Nevertheless, we showed in the previous section that the sign of this expression can be either positive or negative under a more conventional normality assumption based on the primitive preferences of individuals. In this case, it is also true that nothing rules out the possibility for $f_{i1} > 1$. As noted previously, the two normality assumptions are equivalent in the pure public good model, but not with impure altruism. Part of the contribution here, therefore, is to show that the standard normality assumption for impure altruism is in fact more restrictive on preferences than a simple carryover of the same assumption for pure altruism.

Turning now to a change G_{-i}, it holds by definition that $f_{i1} + f_{i2} = \partial G^*/\partial G_{-i}$. To solve for this, we need only add 1 to both sides of (10) and rearrange to find

$$f_{i1} + f_{i2} = \frac{\partial G^*}{\partial G_{-i}} = \pi_G \frac{\partial G^*}{\partial w_i} + (\bar{G}_{\pi_g} - \bar{G}_{\pi_G}) \frac{1}{\Omega}. \quad (12)$$

The standard assumption is that $0 < f_{i1} + f_{i2} < 1$, and as noted in Section 2, this ensures crowding out that is less than one-for-one. While the normality assumption is sufficient to establish this result with pure altruism, equation (12) shows that more is needed to obtain the same bounds on $\partial G^*/\partial G_{-i}$ with impure altruism. In fact, neither of the two different normality assumptions previously discussed is sufficient. The expression is positive if G and g_i are substitutes, can take either sign if they are complements, and nothing rules out the possibility for $f_{i1} + f_{i2} > 1$. Note that (12) greater than 1 implies unfunded crowding in, and (12) negative implies greater than one-for-one crowding out.[17]

Finally, and more important, consider the term f_{i2} on its own, which we can solve for explicitly as the difference between the effect of a change in G_{-i} in (12) and a change in w_i

[17] Ribar and Wilhelm (2002) also make this observation in their footnote 5, where they acknowledge how Cornes and Sandler's (1994) analysis can be used to imply crowding in or greater than one-for-one crowding out with impure altruism.

in (9):

$$f_{i2} = \frac{\partial G^*}{\partial G_{-i}} - \frac{\partial G^*}{\partial w_i} = (\bar{G}_{\pi_g} - \bar{G}_{\pi_G})\frac{1}{\Omega} - \pi_g \frac{\partial G^*}{\partial w_i}. \tag{13}$$

In general, this expression can be positive or negative with either complements or substitutes, and both signs are possible with either of the two normality assumptions. To see how income effects need not determine the sign of this expression, consider the limiting case of quasilinear preferences of the form $x_i + F(G, g_i)$. Without any income effect, such that $\partial G^*/\partial w_i = 0$ or something sufficiently small, $f_{i2} < 0$ requires only net complements with $\bar{G}_{\pi_g} < \bar{G}_{\pi_G}$, otherwise the expression will be positive.[18]

Importantly, the possibility for $f_{i2} < 0$ is inconsistent with the foundation for empirical specification tests between pure and impure altruism, as shown previously in the discussion of (5) and (6). In particular, normality assumptions do not in fact imply $f_{i2} > 0$, and hence crowding out need not be less with impure altruism than with pure altruism. The standard analysis appears to rely on the implicit assumptions of additive separability and strict concavity of the utility function, which is a special case where all of the standard results hold.[19] More generally, however, different possibilities emerge even under very reasonable conditions. Indeed, we find that the relative magnitude of crowding out can be the reverse of conventional wisdom even if $f_{i2} > 0$, because without additive separability, nothing requires f_{i1} to remain constant when comparing impure altruism to the special case of pure altruism. We include several numerical examples in the Appendix that show different possibilities. In particular, we show cases where f_{i2} can take either sign while maintaining $0 < f_{i1} + f_{i2} < 1$, and in all of the cases shown, crowding out is greater with impure altruism than pure altruism.

Table 1 summarizes our general, theoretical results and their relation to the standard assumptions in previous analyses of impure altruism. We find that very few of the standard conditions hold more generally. We have shown that a normality assumption based on preferences, and comparable to that in the pure public good model, is not sufficient to ensure crowding out, except when the crowding effect is funded and G and g_i are net substitutes. If they are net complements, not one of the results has a clear sign or bounded magnitude.

We further summarize our key findings as they relate to empirical work that seeks to

[18] An intuition for $f_{i2} < 0$ begins by recognizing that an increase in G_{-i} makes G less scarce. This implies a decreases in the virtual price of G and an increase in the virtual price of g_i (see the Appendix). The own-price effect pushes for greater demand for G, while the cross-price effect pushes for less if they are complements. By assumption, the cross-price effect is bigger (i.e., $\bar{G}_{\pi_g} < \bar{G}_{\pi_G}$), and demand for G declines despite the increase in G_{-i}. That is, we have greater than one-for-one crowding out with impure altruism even without income effects.

[19] In this case, after differentiating the first order conditions to the utility maximization problem in (2), it straightforward to verify that $f_{i1} = U_{xx}/\Theta$ and $f_{i2} = U_{gg}/\Theta$, where $\Theta = U_{xx} + U_{gg} + U_{GG}$. Hence, with strict concavity, it follows immediately that $0 < f_{i1} < 1$, $f_{i2} > 0$, and $0 < f_{i1} + f_{i2} < 1$. The numerical examples used in previous studies satisfy these conditions (e.g., Andreoni 1990; Ribar and Wilhelm 2002; and Ottoni-Wilhelm, Vesterlund and Xie 2017).

distinguish between pure and impure altruism as a motivation for private provision of a public good. As noted previously, the sizable literature focused on this question assumes that crowding out for impure altruism must be less than that for pure altruism. Nevertheless, we have shown the following:

Result 1 *Both unfunded and funded crowding out with impure altruism can, in general, be more or less than crowding out with pure altruism.*

While this identifies a limitation with the standard specification test used in the literature, our analysis is constructive in the sense that it points to an alternative:

Result 2 *Assuming normality in the usual sense based on preferences means that evidence of crowding in is consistent with impure altruism (i.e., provision of an impure public good), but not with pure altruism.*

The remainder of the paper focuses on empirically estimating the crowding effect in a setting where it is plausible to assume *ex ante* that the model of impure altruism applies. While a finding of either crowding out or in is consistent with impure altruism, a finding of crowding in enables rejection of pure altruism.

5 Empirical Setting and Data Collection

This section begins with institutional background about the Volunteers-in-Parks (VIP) program of the National Park Service (NPS). We then discuss how VIP participation is consistent with the impure altruism motivation for private provision of a public good. This establishes a basis for interpreting our subsequent estimates of a crowding effect, which are based on how changes in budget appropriations to a national park affect the amount of within-park volunteerism. In this section, we also describe our data and report summary statistics.

5.1 The NPS and VIP Program

The NPS is an administrative branch of the U.S. Department of the Interior. The mission of the NPS is to preserve natural and historical landmarks for the enjoyment and education of current and future generations. The NPS system includes over 400 sites, comprises over 84 million acres, and hosts more than 330 million visitors per year. NPS sites include the iconic National Parks that prioritize environmental conservation and outdoor recreation, in addition to parks that emphasize the conservation and restoration of cultural heritage. Examples of the former include Yellowstone and Yosemite National Parks, and examples of the latter include the Statue of Liberty and the Booker T. Washington National Monuments. Throughout the paper, we use the term national park in reference to any of the NPS managed sites.[20] The NPS itself is divided into seven regional offices that manage parks and programs within their geographic jurisdiction, including partial oversight of the VIP program.[21]

Initiated in 1970, the VIP program is a NPS-wide program that facilitates the active involvement of volunteers in protecting and maintaining national parks. Volunteers through the VIP program are an integral part of the NPS workforce, as exemplified by the fact that nearly 340,000 volunteers contributed over eight million hours of service to the national parks in 2016. Volunteer recruitment occurs mainly through word-of-mouth and staff referrals, as well as through online volunteer postings on NPS websites that solicit applications for openings that require specialized skills (NPS 2007). Volunteers are employed across all parks and programs in positions that assist the NPS paid staff of approximately 22,000 employees, or that focus on tasks that could otherwise not be accomplished due to funding shortfalls (NPS 2017a). Funding for the VIP program is separate from other park operations and is regionally allocated before being distributed to individual parks.

Federal appropriations constitute the primary source of funding for national parks. The 2018 NPS budget request exceeded $2.2 billion, while projected revenues from visitor fees are roughly ten percent of this amount. Private donations to the NPS constitute only three percent of the operational needs (U.S. DOI 2019). Throughout the process of formulating the annual NPS budget request, each park begins with a baseline operational budget that is adjusted annually to reflect new projects, maintenance, and programmatic needs (Turner and Walker 2006). Individual park funding requests are then aggregated into the overall NPS budget, which is part of the overall Department of the Interior request. As with funding for all federal agencies, the budget request requires Congressional approval, and the enacted amounts typically differ from the agency requests. The budget appropriations committees in the U.S. House and Senate are those that lead the reconciliation process of the budget between the executive and legislative branches. The U.S. Government shutdown at the end of 2018 and beginning of 2019 illustrates the importance of public funding for park quality. When public funding for parks ceased during the shutdown, parks experienced widespread problems with vandalism, waste, litter, and damage to natural and cultural resources due to unauthorized access of sensitive areas (NPCA 2019).

5.2 Volunteerism and Joint Production

Fundamental to the design of the VIP program is that volunteers can contribute to a public good while simultaneously benefiting from time spent in national parks. The promotional

[20]NPS sites are technically classified into more than two dozen categories depending on their mandate and the types of programs they undertake, including, for example, National Parks, National Monuments, National Historic Sites, and National Battlefields (Comay, 2013).

[21]The seven NPS regions are Alaska, Intermountain, Midwest, National Capital, Northeast, Pacific West, and Southeast.

materials make this explicit, as "the primary purpose of the VIP Program is to provide a vehicle through which the NPS can accept and utilize voluntary help and services from the public in such a way that is mutually beneficial to the NPS and the volunteer" (NPS 2017a). The mutual benefits arise because participation in the VIP program is associated with joint production of public and private goods: the promotion of conservation and opportunities for *in situ* park enjoyment. Volunteerism in the NPS therefore closely matches the notion of private provision of an impure public good (Cornes and Sandler 1994, 1996), which nests the model of impure altruism.[22]

To see how the decision to volunteer in a national park links directly to the model of impure altruism, we need only modify the budget constraint. The composite private good, with a normalized price, must satisfy $x_i = \iota_i + \omega(\tau - v_i)$, where ι_i is the individual's wealth endowment, ω is the wage rate applied to a time budget τ, and v_i is the number of hours spent volunteering. Then, letting $w_i = \iota_i + \omega\tau$ and $g_i = \omega v_i$, we recover the original budget constraint of $x_i + g_i = w_i$, where g_i represents the monetary equivalent of time spent volunteering. Nothing else needs changing about the model other than a scaling for crowding effects depending on the unit of measurement. Specifically, because $\partial g_i^* / \partial G_{-i} = \omega \partial v_i^* / \partial G_{-i}$, the crowding effect on volunteer hours need only be multiplied by the wage rate to have the standard interpretation. The NPS itself places a value on volunteer time using the average, annual hourly wage of non-agricultural workers from the Bureau of Labor Statistics, plus an adjustment of 12 percent to account for fringe benefits. This is the standard approach for estimating a dollar value on volunteer time.[23] During the time period of our study, the average annual value of volunteer time is estimated at $21.85, reported in 2013 dollars.[24]

In this empirical setting, where we study the effect of federal appropriations on volunteerism, there is the question of whether the crowding effects should be interpreted as funded or unfunded. We contend that either is possible because any expected difference between them should be exceedingly small. While experimental studies are able to associate changes in G_{-i} with changes in lump-sum taxation, we examine the effect of changes in G_{-i} without explicit reference to its funding on the part of volunteers. While this perspective suggests an unfunded effect, it must also be recognized that domestic volunteers are taxpayers, so changes in aggregate appropriations do not go unfunded. Moreover, even for volunteers who are not taxpayers, we would argue that their own share of the

[22] If, for example, the private benefit were a warm glow from volunteerism rather than park enjoyment, the impure public good reduces to impure altruism. Chan and Kotchen (2014) generalize the impure public good framework to account for the joint production of multiple public and private goods, which could include, for example, both park enjoyment and a warm glow from volunteerism. Having multiple private benefits does not, however, affect the range of possible theoretical results presented here.

[23] The Independent Sector, a national membership organization of nonprofits, foundations, and corporations, provides an overview of the estimates and their use at https://independentsector.org/value-volunteer-time-methodology/.

[24] All monetary values throughout the paper are reported in 2013 dollars unless otherwise indicated.

required funding would be exceedingly small. This means that any income effects would be correspondingly small, in which case the unfunded and funded crowding effects converge. This is readily seen in the limiting case of quasi-linear preferences, for which the crowding effects in equations (5) and (6) are identical.

5.3 Data

Our primary source of data is a unique and detailed administrative data set from the NPS on annual participation in the VIP program from 1998 through 2013.[25] These data include the annual total number of volunteer hours in each park broken down by the type of volunteer activity. We also obtained data from the NPS on the annual number of full time equivalent (FTE) paid staff in each park. We use publicly available, park-specific data on annual park visitation (1998-2013) and the federal budget appropriation to each park in each year (1995-2013).[26] We collect the League of Conservation Voters (LCV) score for each member of the U.S. House and Senate Appropriations Committees, along with the annual average for both chambers of Congress. The LCV annually scores each voting member of the House and Senate on a 0 to 100 scale based on the percent of pro-environmental and conservation legislation that each member supports.[27] We use the LCV scores to create an instrumental variable for park funding, as we describe in the next section.

Our final sample for analysis includes 326 parks among the 398 originally listed in the VIP data set. The smaller number of parks is due primarily to the way that we include only those that track annual visitation.[28] Figure 3 shows the geographic distribution of the national parks included in our analysis, along with an indicator for each park's NPS administrative region. The first column of Table 2 reports summary statistics across all 326 parks and all years. On average, parks benefit from nearly 16,000 hours of volunteerism per year with an annual value of approximately $346,000, which is roughly nine percent of the average, annual park budget of approximately $3.85 million. Parks host an average of 822,000 visitors per year and employ 0.33 FTE per thousand visitors. The large standard

[25]Data from 2014 through 2018, which we obtained more recently through a Freedom of Information Act request, were collected and categorized into volunteer activities through a different process and unfortunately do not have continuity with the earlier period. We exclude these data from our analysis for this reason, along with the fact that the NPS experienced unique circumstances related to visitation, volunteerism, and management during and around its centennial celebrations in 2016.

[26]Annual park visitation is available online at https://irma.nps.gov/Stats/. Annual park budget appropriations are reported in the U.S. Department of the Interior Budget Justifications for the National Park Service, referred to as the Greenbooks. Selected years are available online at https://www.nps.gov/aboutus/budget.htm.

[27]All LCV data is available online at https://www.lcv.org/. The website also provides a detailed description of the methodology used to generate the scores for each voting member of the U.S. Congress.

[28]Visitation counts are not recorded at 64 smaller urban parks and scenic trails, where unpredictable flows of pedestrian traffic and the existence of multiple access points make consistent counts infeasible. Additionally, we exclude two parks that do not have volunteer data, three parks that do not have their own budgets, one park that does not have information on the number of paid staff, and two parks for which we have only one year of data.

deviations across variables indicate a large degree of heterogeneity among parks.

During the time period of our study, there is significant variation in volunteer hours and funding. Figure 4 shows the annual trends in volunteer hours and federal funding aggregated across all parks. While the total number of volunteer hours has maintained an upward trend, federal funding to the parks follows an upward trend until the recession in 2009. The aggregate trends in Figure 4 do not show a clear pattern in the relationship between park funding and volunteerism, yet our analysis focuses on estimating a causal effect of funding on volunteerism within parks (i.e., not in aggregate). Because of the substantial drop in post-recession funding, we conduct some of our analysis with and without the post-recession years to test for robustness.

One of the distinguishing features of the data set is that we can observe the specific type of activities that volunteers undertake. Figure 5 shows the distribution of volunteer hours across the different types of activities. The most common activity is interpretation, and when combined with natural resource management and maintenance, the three categories comprise over 75 percent of all volunteer hours. Other categories include camp hosting, cultural resource management, and administration, among others.

We use the natural resource management category for the additional purpose of categorizing parks as primarily environmental or non-environmental in some of our analyses that examine heterogeneous effects. We conjecture that volunteering in parks and for activities with more of a natural resource focus are more likely associated with joint production and therefore impure altruism. The rationale is that environmental parks and programs typically focus on recreation, which provide different private benefits to volunteers than parks that focus on curating, for example.[29] We therefore distinguish between environmental and non-environmental parks based on whether or not volunteer hours dedicated to natural resource management are strictly positive for all years within a park. If yes, we classify the park as environmental. This procedure yields 105 environmental parks, and the last two columns of Table 2 report descriptive statistics separately for the two groups. On average, environmental parks have more volunteer hours, greater funding, more visits, and fewer paid FTE per visitor. Later in the paper, we also use an alternative, more general classification based on official park mandates. We also categorize volunteer hours based on whether they are predicted to occur outside or inside in order to examine heterogeneity by volunteer hour type.

[29]For specific examples about how joint production through volunteerism is likely to be greater in environmental parks and outdoor hours, and therefore more readily admit the possibility for crowding in, consider the following: on-going restoration of the Mariposa Grove in Yosemite National Park provides volunteers with additional hiking trails and boardwalks to enjoy while volunteering, whereas federally funded maintenance of the copper skin of the Statue of Liberty is less likely to materially improve the recreational opportunities available to volunteers that provide tours.

6 Empirical Strategy

We now turn to our econometric strategy for estimating the crowding effect and ultimately illustrating our specification test. We focus on estimating the causal effect of changes in federal appropriations to a national park on the amount of within-park volunteerism. We have shown previously that impure altruism admits the possibility for either crowding out or in, and an estimate of crowding in would enable rejection of pure altruism as the underlying motivation for VIP participation. We first describe the benchmark specifications, followed by tests for heterogeneous effects, and our preferred instrumental variables strategy.

6.1 Benchmark Specifications

We begin with fixed effects models of the form

$$Hours_{it} = \beta Budget_{it} + \gamma X_{it} + \alpha_i + \sigma_{rt} + \varepsilon_{it}, \tag{14}$$

where the dependent variable $Hours_{it}$ is the total number of volunteer hours in park i and year t; $Budget_{it}$ is a park's annual budget appropriation in thousands of dollars; X_{it} is a vector of time-varying and park-specific variables; α_i is a vector of park fixed effects; σ_{rt} is a set of year-specific intercepts for each of the NPS seven regions; and ε_{it} is an error term. The variables included in X_{it} are annual park visitation and FTE per 1,000 visits. Standard errors are clustered at the park level in all models to make statistical inference robust to potential serial correlation within parks.

The coefficient of interest is β because it provides an estimate of the crowding effect: a positive estimate indicates crowding in, whereas a negative estimate indicates crowding out. We use volunteer hours as the dependent variable rather than the value of volunteer hours. The reasons are that hours are the original measure of volunteerism, that the conversion to a dollar value requires additional assumptions, and that we can readily obtain a dollar-for-dollar interpretation as an *ex post* adjustment to β, as we will show.

The identifying variation for the crowding effect comes from within-park fluctuations in the annual budget. The park fixed effects capture time-invariant park characteristics, such as popularity, location, type, and the scope of programs, which could affect both funding and volunteer hours. Inclusion of annual visitation controls for changes in a park's popularity over time, due perhaps to anniversary years and promotions that could simultaneously affect funding, volunteerism, and visitation. The inclusion of paid FTE per 1,000 visitors is intended to control for the way that changes in supervisory constraints within a park could potentially affect volunteerism. A positive or negative sign of this effect is also consistent with volunteer hours and permanent FTE serving as complements or substitutes, respectively. The region-year fixed effects control for any annual shocks that are common to all parks within each NPS region.

6.2 Heterogeneous Effects

We next examine heterogeneous effects. We begin by testing whether the crowding effect differs between environmental and non-environmental parks. The estimating equation differs only by the inclusion of an interaction term between $Budget_{it}$ and an indicator variable $1[Envr]_i$ that takes the value of 1 for environmental parks. The full model is

$$Hours_{it} = \beta Budget_{it} + \lambda Budget_{it} \times 1[Envr]_i + \gamma X_{it} + \alpha_i + \sigma_{rt} + \varepsilon_{it}. \tag{15}$$

In this case, β and $\beta + \lambda$ provide estimates of the crowding effect in non-environmental and environmental parks, respectively. A test of whether $\lambda \neq 0$ indicates whether the effect differs between the two types of parks. Testing for the difference is of interest because of the conjecture that volunteerism in an environmental park is more likely to be associated with private provision of an impure public good. Empirical evidence of $\lambda \neq 0$ would therefore support the notion that individual behavior and therefore crowding effects differ in the extent of impure altruism. A finding of crowding in, in either case, would also enable rejection of pure altruism as the underlying motivation for volunteerism in either type of park.

We further examine heterogeneous effects in two ways. First, we examine how crowding effects differ between volunteer activities that are likely to occur outside or inside. Here again the motivating assumption is that volunteer hours focused on outside activities are more likely associated with joint production involving park enjoyment than those taking place inside. For example, the experiences of volunteers who lead guided hiking tours is different from those of volunteers who assist with office administration. Specifically, we estimate the benchmark specification (14), but replace aggregate volunteer hours as the dependent variable with volunteer hours that take place either outside or inside, and estimate the equations separately.[30] Second, we estimate the heterogeneous effects specified in equation (15) with outdoor or indoor hours separately as the dependent variable. This combines the two previous approaches to examine whether the crowding effect differs both by park type (environmental *vs.* non-environmental) and by hour type (outside *vs.* inside).

[30]Outside hours consist of the categories archeology, campground hosting, interpretation, resource management, and park protection. Inside hours consist of the categories general management, curating, administration, maintenance, training, and other.

6.3 IV Strategy

Even after including the park covariates and region-year fixed effects, one might be concerned about potential endogeneity that would bias the crowding effect estimates in equations (14) and (15). The concern would be centered on scenarios where park funding and volunteer hours might be correlated with some unobserved, time-varying park characteristic. For example, a downward trend in park funding that increases the park's deferred maintenance backlog might also increase the demand for park volunteers (NPS 2019b). This would bias the estimates downward because unobserved maintenance needs are positively correlated with volunteerism and negatively correlated with funding. Other plausible scenarios might bias estimates of the crowding effect in the other direction. For example, the unobserved introduction of new park programs could simultaneously create additional volunteer opportunities and funding requirements.

Addressing concerns about endogeneity requires an instrument that is correlated with the annual budget for each park, but uncorrelated with time-varying volunteerism within each park, conditional on the model covariates. We construct an instrument that plausibly meets these two requirements by combining cross-sectional variation in historical budgetary needs with shifts to public funding in a manner that is commonly used in the literature on trade and local labor markets (Bartik 1991; Autor, Dorn and Hanson, 2013; Notowidigdo 2019). We simulate annual park funding in each year with the interaction between each park's funding levels three years prior to the start of our panel and changes in the average LCV score for the U.S. Congressional Appropriations Committees, excluding members that represent the corresponding park. We use the House and Senate appropriation committees because their members have the most significant influence on federal budget appropriations each year. For each park, we exclude any members of Congress representing their district or state to avoid the possibility that factors affecting volunteerism in a park might also affect the preferences of elected officials, and thereby violate the exclusion restriction.

Specifically, the instrument is defined as

$$Z_{it} = Budget_{i,1995} \times \frac{\overline{LCV}_{i,t-1}}{\overline{LCV}_{i,1995}}, \tag{16}$$

where $Budget_{i,1995}$ is each park's federal budget appropriation in 1995, and $\overline{LCV}_{i,t}$ is the average of the LCV scores of the two committees in year t, excluding the elected official(s) of park i.[31] The $Budget_{i,1995}$ weights capture differences in each park's base budgetary needs due to their size or historical program scope. The ratio $\overline{LCV}_{i,t-1}/\overline{LCV}_{i,1995}$ captures

[31] In particular, we estimate the average among members in the House and Senate committees separately, and then take the average of the two committees to estimate $\overline{LCV}_{i,t}$ for each year. In any given year, the appropriations committees consist of approximately 60 members in the House and 30 members in the Senate.

the year-to-year change in the extent to which the key congressional committees prioritize environmental and conservation legislation.[32] The IV strategy is therefore based on the theory that appropriations committees more supportive of these objectives will allocate higher levels of funding to the NPS. We lag the LCV score because the budget approved in year $t-1$ is for use in year t. Figure 6 shows the time series variation of the mean LCV score, and referring back to Figure 4, the trend appears to roughly track the variation in NPS funding.

We use Z_{it} to instrument for $Budget_{it}$ in specifications (14) and (15), and $Z_{it} \times 1[Envr]_{it}$ as an instrument for $Budget_{it} \times 1[Envr]_i$ in specification (15).[33] Table 3 reports the first stage results, and we find that the instrument is highly correlated with actual park funding.[34] Column 1 shows that a 10 percent increase in LCV score increases park budgets by 2 percent relative to their budget in 1995.

Note that the approach is strengthened by the way that all models include region-year fixed effects. Controlling for these, the identifying variation comes from the way that exogenous changes in aggregate LCV scores scale each individual park's budget, where additional heterogeneity comes from differences in the level of each park's 1995 budget. This means that, in any given year, the assumption of instrumental exogeneity would fail only if volunteers systematically redirect their efforts to different parks within a region in response to changes in the LCV score. While such a response seems unlikely to us, there is also evidence that 78.4 percent of the NPS volunteers report their primary motivation as interest in specific parks and projects, along with the overall NPS mission (NPS 2007).

In the next section, as part of robustness checks, we discuss and use alternative instruments to scale each park's initial funding by other variables that predict aggregate NPS funding as in (16). These include LCV scores for the all members of the U.S. House and Senate (rather than just the appropriation committees), annual funding for the Department of the Interior's Fish and Wildlife Service, and the annual NPS funding for all regions excluding the region for each park.

[32] We choose the base year 1995 because it is three years before the start of our VIP data and is the earliest year for which we observe funding. While the choice of other possible base years changes the coefficient estimates in the first stage, it has no effect on the second stage IV estimates.

[33] There are 36 parks in our sample that do not exist in 1995, so there is no $Budget_{i,1995}$ for these observations. In these cases, we use the first year of budget data available, and include observations beginning three years after the designation of these parks. Excluding or including these observations does not change any of the results.

[34] The cluster-robust Kleibergen-Paap rk Wald F-statistic for a weak instrument shows that while the instrument is highly correlated with park funding across all parks, it is less correlated with the non-environmental parks (column 2). Relative to the benchmark specification (14), this produces a lower test statistic for joint identification in the two first-stage regressions for specification (15) ($F = 6.4$). This means that the heterogeneous IV estimates of specification (15) may be more biased toward OLS.

7 Estimation Results

We now turn to our estimates of the effect of changes in park funding on within-park volunteerism. We consider the overall, average estimates of the crowding effect, heterogeneity between park and volunteer hour types, a range of robustness checks, and alternative explanations.

7.1 Crowding Out or In?

The estimates in Table 4 address the question of whether increases in park funding crowd out or crowd in volunteerism on average across parks. The first three columns report the OLS estimates of specification (14). Model (1) includes only $Budget_{it}$ along with the park and region-year fixed effects, model (2) includes the additional variables of annual visits and FTE per 1,000 visits, and model (3) is the same though estimated excluding the post-recession years 2011-2013. We find positive and statistically significant coefficients on park budgets across all three models, and the estimated magnitudes are very similar. Recall that the positive coefficients are consistent with crowding in, rather than crowding out. The OLS estimates suggest that a \$1,000 increase in a park's annual budget is associated with an increase in volunteerism of approximately 2.5 hours on average.

Our preferred estimates of the average crowding effect are the results of the IV estimation in the last three columns of Table 4. These models correct for potential endogeneity of park budgets using the IV strategy described above. The three models differ in parallel fashion with those in the first three columns. Here again the results are consistent with crowding in, and the magnitudes are larger and more precisely estimated. Focusing on the results of model (5), which includes the covariates, we find that a \$1,000 increase in federal funding causes an average increase in volunteerism of approximately 12.4 hours. We also find that the results are economically meaningful. To quantify the average value of crowding in, we multiply the estimated change in hours by the average hourly valuation of volunteerism over the years 1998 through 2013. As described previously, this estimate is \$21.85 per hour. Accordingly, based on our preferred model (5), a \$1,000 increase a park's federal appropriation crowds in volunteerism that is worth \$271 per year on average. This means that an additional dollar of federal funding within a park crowds in roughly 27 cents worth of volunteerism, where the benefit of crowding in occurs over and above the direct benefit of the marginal dollar of park funding. As noted previously, one reason why the magnitudes of the IV estimates might be larger than the OLS estimates is that lower levels of funding add to the cumulative deferred maintenance in parks, and this increases the set of potential tasks for volunteers.[35]

[35] In 2017, the deferred maintenance backlog in the NPS was estimated at \$11.61 billion in current dollars (NPS 2019b).

The number of visits has no statistically significant effect on volunteerism, but the FTE per 1,000 visits does. With greater paid staff per visitor, there is less volunteerism, and the magnitude is such that one additional FTE is associated with roughly 50 fewer hours of volunteerism. The sign of this effect is important because it suggests that paid staff and volunteerism are substitutes rather than complements when it comes to park management. This is a subject to which we return below, but it is worth noting here that the finding helps to rule out an alternative explanation for crowding in, whereby greater budgets result in more paid staff, who are then able to recruit and manage more volunteers.[36]

In sum, we find evidence consistent with a mechanism whereby greater funding results in better parks that are more enjoyable places in which to volunteer. We have shown how the presence of such joint production readily admits the possibility for crowding in, whereas the model for private provision of a pure public good does not. We therefore interpret the results of crowding in, along with the differences between the OLS and IV estimates, as consistent with private provision of an impure public good (i.e., impure altruism), and we reject the model of a pure public good, which only admits crowding out.

7.2 Heterogeneous Effects

We now test for heterogeneous effects of our crowding estimates across park and hour types. We begin with potential differences between parks designated as either environmental or non-environmental. Recall our conjecture that the joint production of volunteerism is greater in the environmental parks, where individuals are more likely to experience *in situ* benefits of park enjoyment (i.e., recreating in a natural environment) while contributing to the public good.

Table 5 reports the estimates of specification (15), and we again show the OLS and IV results. In all models, we find positive and statistically significant coefficient estimates on the interaction $Budget_{it} \times 1[Envr]_i$, which indicates more crowding in for environmental parks than for non-environmental parks. Focusing on the IV estimates, particularly those for model (5), the magnitude of the difference indicates that a \$1,000 increase in park funding crowds in 11.6 more hours in the environmental parks on average. The overall effect in environmental parks, which is the sum of the first two coefficients, is 12.9 hours, which has an equivalent monetary value of \$282. That is, each additional dollar of park funding crowds in roughly 28 cents worth of volunteerism. In contrast, the crowding effect in non-environmental parks is statistically indistinguishable from zero. Based on our theoretical results for distinguishing between models based on a finding of crowding in, we conclude that volunteerism in environmental parks is consistent with private provision of an impure

[36]The substitutability, rather than complementarity, between paid staff and volunteers is also echoed in survey responses as part of the 2007 VIP Program Assessment Report, where one staff member states in response to budget shortfalls that, "In my 16-plus years, I have seen the volunteer role evolve to replace many of the functions that National Park Service staff previously performed" (NPS 2007).

public good, but not a pure public good. However, no distinction between the models is possible for volunteerism in non-environmental parks.

Turning now to heterogeneous effects by hour type, Table 6 reports estimates of specification (14) separately for outside and inside hours in panels A and B, respectively.[37] Focusing on the IV estimates, we find evidence of crowding in for hours of both types, although the magnitudes and precision are greater for outside hours. Based on the preferred models in column (4), we find crowding in of 9.3 outside hours, compared to crowding in of 3.1 inside hours. These translate to monetary equivalent measures of 20 cents and 7 cents worth of volunteerism per dollar of federal budgeting, respectively. The difference between the two estimates—and the fact that both indicate crowding in—remains consistent with the general notion that environmental parks and outdoor hours create circumstances where and when joint production of volunteerism is likely to be greatest. A further difference worth noting between outdoor and indoor hours is the extent of substitutability between volunteerism and paid FTE. One additional full-time paid staff member per 1,000 visitors is associated with 71 fewer volunteer hours outside, but has no statistically significant effect on volunteer hours inside.

The final set of heterogeneous effects that we examine is a two-way analysis, where we estimate the environmental versus non-environmental park effect separately for outside and inside hours. Table 7 reports estimates of specification (15) for outside and inside hours separately in panels A and B, respectively. Focusing on the preferred estimates in column (4), we find evidence that environmental parks differ from non-environmental in the direction of more crowding in for both outside and inside hours. Moreover, as expected, the magnitudes of the point estimates suggest greater crowding in for outside hours than for inside hours, at 7.7 versus 3.9 hours per $1,000 of funding. Nevertheless, among the four different cases of the two-way analysis, the overall estimate of crowding in is statistically significant at conventional levels in only one case: outside hours in environmental parks, with crowding in of 9.6 hours per $1,000 of funding, and a 95 percent confidence interval that ranges between 1.6 and 17.7 hours.

7.3 Robustness Checks

We estimate a series of alternative models to demonstrate robustness of our results on crowding in and heterogeneity. In particular, we consider lags on selected explanatory variables, the addition of a linear time trend that differs between large and small parks, a broader definition of environmental parks, and alternative instruments in our IV strategy.[38] We report the full set of results in Appendix Tables A.2 and A.3 and summarize the

[37]We no longer report estimates that exclude the post-recession years for two reasons: the results do not change in any meaningful way, and we prefer to focus on results that take advantage of the full data set.

[38]Note that the results reported earlier where we exclude the post-recession years (2011-2013) should also be considered robustness checks. We found that dropping these years had no meaningful affect on the results.

key findings here.

We first consider a one-year lag of the instrumented annual park budgets. Central to our analysis is the idea that greater funding improves parks in ways that may affect volunteerism, but many funded initiatives take time to complete. Accordingly, one could argue that funding in previous years is a preferable measure of improvements in park quality, and we therefore estimate specifications (14) and (15) using an instrumented version of the lagged budget, $Budget_{i,t-1}$, on the right-hand side rather than the contemporaneous year's budget. One consequence of using the lagged variable is having fewer observations upon which to estimate the model. Nevertheless, we again find statistically significant crowding in, and the magnitudes are greater than those estimated previously (Table A.2 columns 1 and 2). For the overall effect across parks, we find crowding in of 13.9 hours per $1,000 of funding on average, or equivalently 30 cents worth of volunteerism for an additional dollar of funding. The estimated effect for environmental parks is quite similar in sign, magnitude and precision to the overall estimate of crowding in, and we again find no statistically significant crowding effect in non-environmental parks.

We also estimate models that include the additional controls of one- and two-year lags of park visitation. Contemporaneous visitation is indicative of a park's popularity in a given year, but visitation from prior years is the only information available to Congress when the NPS budget must be approved. Accounting for the popularity of parks in prior years might also capture how popularity affects volunteerism, which in many cases may require planning far in advance. It turns out that the visitation variables themselves have statistically insignificant effects, and we find that the estimates of crowding in remain very similar to those reported earlier for specifications (see Table A.2, columns 3 and 4).

A further robustness check is to include separate linear time trends for parks with budgets above and below the median in 1995. This accounts for the possibility that volunteerism might be affected by different trends over time in large and small parks. Here again we find that the results are very similar to those already reported (see Table A.2 columns 5 and 6).

Our tests for heterogeneous effects based on park type relied on a particular definition for environmental parks (i.e., whether the park engages volunteers in natural resource management every year). This definition yields 105 environmental parks that, on average, had more volunteers, larger budgets, greater visitation, and fewer FTE per visitor (see Table 2). As a robustness check, we consider an alternative definition of an environmental park based on each park's official mandate. We reviewed the descriptions of each park's activities on its official webpage and classified parks as environmental if they list natural resource management as part of their mission, regardless of whether conservation is a primary or secondary objective.[39] The result is a broader definition of environmental parks,

[39]For example, National Battlefields, such as Antietam National Battlefield, are established with the primary

accounting for 233 of all 326 parks, and nesting 99 of the 105 included in the previous definition. Table A.1 includes the descriptive statistics for parks of both types based on this alternative definition. We find that testing for heterogeneous effects using this definition of environmental parks in specification (15) results in a similar estimate of crowding in for environmental parks, at 12.5 hours compared to 11.6 hours previously, and still no statistically significant effect in non-environmental parks (Table A.2 column 5).

Finally, we consider alternative instruments, while still using the basic shift-share approach in equation (16). All results are reported in Table A.3, though we note that our preferred instrument in a better predictor of funding in the heterogeneous effects models. In each case, we scale each park's base year funding with the change in a variable that predicts aggregate changes in NPS funding. First, we use the annual shift in the LCV score for the entire U.S. Congress, averaged across the whole House and Senate, rather than only the appropriations committees. In this case, we continue to exclude the members representing the corresponding park. We find very similar estimates of crowding in for both specifications (14) and (15). Second, in place of LCV scores, we use the shift in the annual budget of the U.S. Fish and Wildlife Service, which is another Bureau within the U.S. Department of the Interior with a similar environmental orientation as the NPS.[40] This approach is based on actual funding decisions that are plausibly related to funding levels in the NPS, and we again find results that are similar in sign, magnitude, and precision. Finally, for each park i, we use the total NPS budget in each year, exclusive of funding for the region in which park i is located. With this approach, the scaling of each park's budget differs among the seven NPS regions, and the magnitude of the crowding in estimates are somewhat smaller but have the same qualitative interpretation.

7.4 Alternative Explanations

Might there be alternative explanations for crowding in of volunteerism in the NPS? One possibility for crowding in that has been raised in the literature more broadly is a signaling explanation. For example, Payne (2001) finds that greater public funding can signal greater governmental approval of a recipient organization, which, in turn, promotes an increase in private contributions. In a related finding, Khanna and Sandler (2000) show how private contributions might be greater for institutions perceived as subject to more stringent governmental oversight. We argue that a signaling explanation is unlikely to explain the crowding in of volunteerism in national parks for several reasons. First, the

goal of commemorating specific historic events. However, Antietam National Battlefield is classified as an environmental park based on its mandate because it additionally records and preserves local wildlife and vegetation, monitors invasive species and other pests, and reforests the land on which it is established (NPS 2019a).

[40] Data for the total budgets for all Bureaus within the Department of the Interior are available in the annual Greenbooks. In this case, the ratio in equation (16) does not differ by park.

inclusion of park visitation in our regression models, both contemporaneous and lagged, controls for quality signaling that would plausibly affect visitors and volunteers in the same way. Second, the park fixed effects control for any persistent quality signal among parks, which is likely to be more relevant given the opacity of annual budgeting. We find it easier to argue that volunteers are more likely to observe park or program quality than annual fluctuations in a park's budget. Finally, all of the parks in our analysis are subject to NPS authority and thereby face a similar set of oversight standards for management and accountability. Indeed, it is the relative homogeneity across our units of analysis that provides a distinct advantage compared to other studies in the literature that seek to estimate crowding effects across a range of heterogeneous charitable organizations.

Another candidate explanation, relates to the possibility of endogenous fundraising, whereby changes in public funding can affect the incentives of charitable organizations to independently solicit donations (Andreoni and Payne 2001, 2003). While fundraising plays a relatively small role in the NPS, concerns might arise in our setting about the ability of park staff to recruit and manage volunteers. In particular, might the estimated crowding in be the result of greater funding relaxing a binding constraint on the ability of park managers to recruit and supervise volunteers? Here again, there are several reasons why we believe this mechanism is not explaining our results. First, we include the paid FTE per visitor as an explanatory variable in the preferred regression models, which controls for annual park-specific variation in staff availability to manage volunteers. In fact, we find robust evidence that more FTE per visitor is negatively associated with the number of volunteer hours within a park. Second, an evaluation of the VIP program provides broad survey evidence that supervisory capacity is not generally a binding constraint (NPS 2007).[41] Third, the VIP program is regionally funded before allocations are made to individuals parks, and this means that the region-year fixed effects would control for any common, regional funding constraints (NPS 2005).[42] Fourth, even the largest parks employ only a handful of designated VIP program administrators, and this suggests that increases in operational funding are unlikely to be used for expanding the VIP program staff within parks (NPS 2017b). Finally, our estimates of heterogeneous effects would mean that supervisory constraints would need to apply differently to environmental and non-environmental parks, along with outside and inside hours within the same park. While possible, we see no compelling reason why such differences would arise.

[41]In particular, the survey finds that 88 percent of the volunteers "do not need more of their supervisor's time" and 89 percent are at least moderately "satisfied with the leadership, management, and support that [they] receive as volunteers," with 77 percent reporting strong satisfaction. Moreover, only 16 percent of the staff agree with the statement that "volunteers are not given the necessary attention and direction throughout their assignments" (NPS 2007).

[42]We have sought to obtain detailed data on park-level VIP budgets in each year, but unfortunately the NPS does not keep track of these data in regional offices.

8 Discussion and Conclusion

This paper makes theoretical and empirical contributions to the literature on private provision of public goods. Understanding the ways in which public provision affects private provision—through either crowding in or crowding out—is fundamental for evaluating the positive and normative consequences of policies that affect the supply of public goods. Moreover, within the literature on privately provided public goods, estimates of the extent of crowding out are frequently used to test between the candidate models of underlying behavior based on pure or impure altruism. This paper identifies limitations of the generality of the standard specification test based on crowding out, proposes a more general alternative test based on crowding in, and provides empirical evidence consistent with the new test in a setting where impure altruism appears to apply *ex ante*, because of joint production of private and public benefits.

Our theoretical analysis revisits the crowding out conditions in Andreoni's (1989, 1990) seminal model of impure altruism. We show that the typically asserted normality assumptions are less straightforward than they might appear. The subtlety arises because of how the model's constrained setup creates a difference between the normality of goods based on primitive preferences and the income effects that shift demand functions. In contrast with the pure public good model, we show that crowding out with impure altruism is an assumed result, rather than a consequence of well-understood properties of individual utility functions. Applying a more general insight of Cornes and Sandler (1984, 1994) to the particular setting of impure altruism, we show how assuming normality in the usual sense based on preferences readily admits possibilities that are implicitly assumed away: crowding in and greater than one-for-one crowding out.

More novel and potentially important results relate to the specification tests between pure and impure altruism. Contrary to the conventional wisdom in the literature that crowding out must be less for impure altruism than for pure altruism, we show that this need not be the case. The standard results will always hold under the assumptions of additive separability and strict concavity of utility functions, but more generally crowding out with impure altruism can be greater than or less than with pure altruism. We show that the different cases depend in part on the degree of substitutability or complementarity between the public good and the private benefit associated with one's own provision. Furthermore, these results do not hinge on different normality assumptions. Indeed, we show examples that illustrate the different possibilities under either version of the normality assumption and even without any income effects. As an alternative, we propose a more general specification test based on crowding in: evidence of crowding in is consistent with impure altruism but not with pure altruism assuming normality of preferences.

The empirical portion of the paper focuses on estimating the causal effect of changes in public funding within U.S. national parks on within-park volunteerism. We find robust evidence of crowding in across a range of specifications and identification strategies, relying on OLS and IV estimates, and heterogeneous effects. The overall finding is that each additional dollar of public expenditure within a park crowds in 27 cents worth of volunteerism on average. Our empirical setting and data set also have unique advantages for estimating crowding effects because of the relatively long panel of homogeneously managed units, volunteers that tend to donate their time at only one site, and federal funds that account for the vast majority of park budgets. From a policy perspective, our findings also suggest that greater budgets for the NPS will meaningfully leverage greater private contributions, and this may be an important consideration for a government agency that suffers from perennial budgetary shortfalls and a growing backlog of deferred maintenance.

Finally, we have argued that volunteerism in national parks, which involves giving time in parks rather than money, is consistent with private provision of an impure public good and therefore the model of impure altruism. The argument is based on how volunteerism jointly contributes to park conservation and management (a public good) and *in situ* park enjoyment (a private benefit). In the context of our theoretical results and new specification test, the findings of crowding in support the notion that volunteerism is driven by the underlying model of impure altruism rather than pure altruism. The empirical findings thus provide well-identified, quasi-experimental evidence of crowding in that is theoretically consistent with a generalized interpretation of the standard model for private provision of a public good.

Appendix

Derivation of Equations (9), (10), and (11)

We follow Cornes and Sandler's (1984, 1994) more general approach for deriving comparative static properties of the impure public good model. There are three conditions that link the virtual magnitudes (π_g, π_G, m_i) to changes in the exogenous parameters. First is the relationship between an individual's own provision and the level of aggregate provision:

$$g_i(\pi_g, \pi_G, m_i) = G(\pi_g, \pi_G, m_i) - G_{-i}. \tag{17}$$

This is simply a restatement of the identity in (4). Second is that the virtual prices of the jointly produced products must sum to the price of the joint product:

$$\pi_g + \pi_G = 1. \tag{18}$$

The third comes from substituting (17) and (18) into the virtual budget constraint, $x_i + \pi_G G + \pi_g g_i = m_i$, along with the actual budget constraint $x_i + g_i = w_i$, and solving for

$$m_i = w_i + \pi_G G_{-i}, \tag{19}$$

which is simply the virtual "full income." Recognizing that each of the virtual magnitudes is a function of the vector of exogenous parameters (w_i, G_{-i}), we can differentiate (17), (18), and (19) with respect to either parameter, denoted by θ, and use Cramer's rule to solve for $\partial\pi_g/\partial\theta$, $\partial\pi_G/\partial\theta$, and $\partial m_i/\partial\theta$. In particular, following this procedure for a change in w_i yields

$$\begin{aligned}
\frac{\partial \pi_g}{\partial w_i} &= \frac{g_{m_i} - G_{m_i}}{\Omega}, \\
\frac{\partial \pi_G}{\partial w_i} &= \frac{G_{m_i} - g_{m_i}}{\Omega}, \\
\frac{\partial m_i}{\partial w_i} &= 1 + \frac{G_{-i}(G_{m_i} - g_{m_i})}{\Omega},
\end{aligned}$$

where as defined in the main text $\Omega = \bar{g}_{\pi_G} + \bar{G}_{\pi_g} - \bar{g}_{\pi_g} - \bar{G}_{\pi_G}$. The parallel results for a change in G_{-i} are

$$\begin{aligned}
\frac{\partial \pi_g}{\partial G_{-i}} &= \frac{\pi_G(g_{m_i} - G_{m_i}) + 1}{\Omega}, \\
\frac{\partial \pi_G}{\partial G_{-i}} &= \frac{\pi_G\left(G_{m_i} - g_{m_i}\right) - 1}{\Omega}, \\
\frac{\partial m_i}{\partial G_{-i}} &= \pi_G + \frac{G_{-i}\left(\pi_G\left(G_{m_i} - g_{m_i}\right) - 1\right)}{\Omega}.
\end{aligned}$$

Finally, substituting these expressions into equation (8), along with the Slutsky equation to decompose the uncompensated prices responses, and rearranging immediately yields equations (9), (10), and (11).

Numerical Examples

We construct numerical examples to show the key result that crowding out can be greater with impure altruism than with pure altruism. Our examples use different variants of the quadratic utility function. In each case, pure altruism is a special case of impure altruism. The exogenous variables are always normalized so that $w_i = G_{-i} = 1$. The parameters are also set to ensure that the quadratic utility function is strictly increasing locally around the optimal solution. For alternative (i.e., higher) numerical values of the exogenous variables, a rescaling of the utility function might be required to ensure the utility function is well-behaved.

We use the superscript alt to denote results for pure altruism, and the superscript imp to denote results for impure altruism. Specifically, we are comparing f_{i1}^{alt} with $f_{i1}^{imp} + f_{i2}^{imp}$ at the chosen consumption bundle, noting that by definition $f_{i2}^{alt} = 0$ and paying special attention the sign of f_{i2}^{imp}. It holds in all cases that $0 < f_{i1}^{alt} < 1$ and $0 < f_{i1}^{imp} + f_{i2}^{imp} < 1$, which implies crowding out. In all three examples $f_{i1}^{alt} > f_{i1}^{imp} + f_{i2}^{imp}$, which means that crowding out is *greater* with impure altruism, and we show how this is possible regardless of whether $f_{i1}^{alt} \gtreqless f_{i1}^{imp}$, or whether $f_{i2}^{imp} \lesseqgtr 0$.

Example 1 *This example shows a case where $f_{i1}^{alt} = f_{i1}^{imp}$ and $f_{i2}^{imp} < 0$. Preferences for pure and impure altruism are*

$$\begin{aligned} U(x_i, G) &= \frac{23}{8}x_i - \frac{x_i^2}{2} + 4G - G^2 + x_i G \\ U(x_i, G, g_i) &= U(x_i, G) - g_i^2 - x_i g_i + \frac{5}{4} g_i G, \end{aligned}$$

and it follows that $f_{i1}^{alt} = \frac{4}{10}$ and $f_{i1}^{imp} + f_{i2}^{imp} = \frac{4}{10} - \frac{1}{10} = \frac{3}{10} < f_{i1}^{alt}$.

Example 2 *This example shows a case where $f_{i1}^{alt} > f_{i1}^{imp}$ and $f_{i2}^{imp} < 0$. Preferences for pure and impure altruism are*

$$\begin{aligned} U(x_i, G) &= 3x_i - x_i^2 + 4G - G^2 + 4x_i G \\ U(x_i, G, g_i) &= U(x_i, G) + 4g_i - g_i^2 - 3x_i g_i, \end{aligned}$$

and it follows that $f_{i1}^{alt} = \frac{1}{2}$ and $f_{i1}^{imp} + f_{i2}^{imp} = \frac{3}{8} - \frac{1}{8} = \frac{1}{4} < f_{i1}^{alt}$.

Example 3 *This example shows a case where $f_{i1}^{alt} > f_{i1}^{imp}$ and $f_{i2}^{imp} > 0$. Preferences for pure and* *impure altruism are*

$$\begin{aligned} U(x_i, G) &= 3x_i - x_i^2 + 4G - G^2 + 4x_i G \\ U(x_i, G, g_i) &= U(x_i, G) + 4g_i - g_i^2 - \frac{3}{2}x_i g_i, \end{aligned}$$

and it follows that $f_{i1}^{alt} = \frac{1}{2}$ *and* $f_{i1}^{imp} + f_{i2}^{imp} = \frac{9}{22} + \frac{1}{22} = \frac{5}{11} < f_{i1}^{alt}$.

References

Andreoni, James. 1989. “Giving with Impure Altruism: Applications to Charity and Ricardian Equivalence." *Journal of Political Economy* 97 (6): 1447-1458.

Andreoni, James. 1990. “Impure Altruism and Donations to Public Goods: A Theory of Warm-Glow Giving.” *Economic Journal*, 100 (401): 464-477.

Andreoni, James. 1993. “An Experimental Test of the Public Goods Crowding-Out Hypothesis.” *American Economic Review* 83 (5): 1317-1327.

Andreoni, James. 2006. “Philanthropy.” In *Handbook of Giving, Reciprocity and Altruism*, Vol. 2, edited by Serge-Christophe Kolm and Jean Mercier Ythier. North-Holland: Elsevier.

Andreoni, James and A. Abigail Payne. 2003. “Do Government Grants to Private Charities Crowd Out Giving or Fund-raising?" *American Economic Review* 93 (3): 792-812.

Andreoni, James and A. Abigail Payne. 2011. “Is Crowding Out Due Entirely to Fundraising? Evidence from a Panel of Charities.” *Journal of Public Economics* 95 (5-6): 334-343.

Andreoni, James and A. Abigail Payne. 2013. “Charitable Giving.” In *Handbook of Public Economics*, Vol. 5, edited by Alan J. Auerbach, Raj Chetty, Martin Feldstein,and Emmanuel Saez. Amsterdam: Elsevier.

Autor, David H., David Dorn, and Gordon H. Hanson. 2013. “The China Syndrome: Local Labor Market Effects of Import Competition in the United States.” *American Economic Review* 103(6): 2121-2168.

Bartik, Timothy. 1991. “Who Benefits from State and Local Economic Development Policies?” W.E. Upjohn Institute for Employment Research: Kalamazoo, MI.

Becker, Gary S. 1974. “A Theory of Social Interaction.” *Journal of Political Economy* 82 (6): 1063-1093.

Bergstrom, Theodore, Lawrence Blume, and Hal Varian. 1986. “On the Private Provision of Public Goods.” *Journal of Public Economics* 29 (1): 25-49.

Borgonovi, Francesca. 2006. “Do Public Grants to American Theaters Crowd-out Private Donations?” *Public Choice* 126 (3/4): 429-451.

Chan, Nathan W. and Matthew J. Kotchen. 2014. “A Generalized Impure Public Good and Linear Characteristics Model of Green Consumption.” *Resource and Energy Economics* 37: 1-16.

Comay, Laura B. 2013. "National Parks System: What Do the Different Park Titles Signify?" Report 41816. Washington, DC: Congressional Research Service.

Cornes, Richard and Todd Sandler. 1984. "Easy Riders, Joint Production, and Public Goods." *Economic Journal* 94 (375): 580-598.

Cornes, Richard and Todd Sandler. 1985. "The Simple Analytics of Pure Public Good Provision," *Economica* 52(205): 103-116.

Cornes, Richard and Todd Sandler. 1994. "The Comparative Static Properties of the Impure Public Good Model." *Journal of Public Economics* 54 (3): 403-421.

de Wit, Arjen and Rene Bekkers. 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.

Eckel, Catherine C., Philip J. Grossman, and Rachel M. Johnston. 2005. "An Experimental Test of the Crowding Out Hypothesis." *Journal of Public Economics* 89 (8): 1543-1560.

Glazer, Amihai and Kai A. Konrad. 1996. "A Signaling Explanation for Charity." *American Economic Review* 86 (4): 1019-1028.

Harbaugh, William T. 1998. "The Prestige Motive for Making Charitable Transfers." *American Economic Review* 88 (2): 277-282.

Hines, James R. and Richard H. Thaler. 1995. "The Flypaper Effect." *Journal of Economic Perspectives* 9 (4): 217-226.

Hollander, Heinz. 1990. "A Social Exchange Approach to Voluntary Cooperation." *American Economic Review* 80 (5): 1157-1167.

Hungerman, Daniel M. 2005. "Are Church and State Substitutes? Evidence from the 1996 Welfare Reform." *Journal of Public Economics* 89 (11-12): 2245-2267.

Independent Sector. 2019. "The Value of Volunteer Time." https://www.independentsector.org /resource/the-value-of-volunteer-time/ (accessed February 25, 2019).

Khanna, Jyoti and Todd Sandler. 2000. "Partners in Giving: The Crowding in Effects of U.K. Government Grants." *European Economic Review* 44 (8): 1543-1556.

Kingma, Bruce R. 1989. "An Accurate Measurement of the Crowd-out Effect, Income Effect, and Price Effect for Charitable Contributions." *Journal of Political Economy* 97 (5): 1197-1207.

Kotchen, Matthew J. 2005. "Impure Public Goods and the Comparative Statics of Environmentally Friendly Consumption." *Journal of Environmental Economics and Management* 49 (2): 281-300.

Kotchen, Matthew J. 2006. "Green Markets and the Private Provision of Public Goods." *Journal of Political Economy* 144 (4): 816-834.

Kotchen, Matthew J. 2007. "Equilibrium Existence and Uniqueness in Impure Public Good Models." *Economic Letters* 97: 91-96.

League of Conservation Voters (LCV). 2019. "National Environmental Scorecard Methodology." http://scorecard.lcv.org/methodology/ (accessed January 16, 2019).

Murdoch, James C. and Todd Sandler. 1984. "Complementarity, Free Riding, and the Military Expenditures of NATO Allies," *Journal of Public Economics*. 25(1-2) 83-101.

Murdoch, James C. and Todd Sandler. 1990. "Nash-Cournot or Lindahl Behavior? An Empirical Test for the NATO Allies," *Quarterly Journal of Economics*. 105(4) 875-894.

National Park Service (NPS). 2005. "Director's Order #7: Volunteers-In-Parks." https://www.nps.gov/policy/DOrders/DO-7.htm/ (accessed May 5, 2017).

National Park Service (NPS). 2007. "VIP Program Assessment Report." United States Department of the Interior, Washington, DC.

National Park Service (NPS). 2017a. "NPS Essentials." https://www.nps.gov/training/essentials /html/volunteers.html (accessed May 6, 2017).

National Park Service (NPS). 2017b. "Yosemite National Park Volunteers in Parks Handbook." https://www.nps.gov/yose/getinvolved/supportyourpark/upload/viphandbook.pdf (accessed May 4, 2017).

National Park Service (NPS). 2019a. "Antietam Nature and Science," https://www.nps.gov/anti/learn/nature/index.htm/ (accessed February 26, 2019).

National Park Service (NPS). 2019b. "What is Deferred Maintenance?" https://www.nps.gov/subjects/infrastructure/deferred-maintenance.htm/ (February 28, 2019).

National Parks Conservation Association (NPCA). 2019. "How Is the Partial Government Shutdown Affecting National Parks?" (accessed on March 11, 2019). https://www.npca.org/articles/2029-how-is-the-partial-government-shutdown-affecting-national-parks

Notowidigdo, Matthew. 2019. "The Incidence of Local Labor Demand Shocks." Forthcoming, *Journal of Labor Economics*.

Okten, Cagla and Weisbrod, Burton A. 2000. "Determinants of Donations in Private Non-Profit Markets." *Journal of Public Economics* 75 (2): 255-272.

Ottoni-Wilhelm, Mark, Lise Vesterlund, and Huan Xie. 2017. "Why Do People Give? Testing Pure and Impure Altruism." *American Economic Review* 107 (11): 3617-3633.

Payne, A. Abigail. 2001. "Measuring the Effect of Federal Research Funding on Private Donations at Research Universities: Is Federal Research Funding More Than a Substitute for Private Donations?" *International Tax and Public Finance* 8 (5-6): 731-51.

Payne, A. Abigail. 2009. "Does Government Funding Change Behavior? An Empirical Analysis of Crowding Out." In *Tax Policy and the Economy*, Vol. 23, edited by Jeffrey R. Brown and James M. Poterba. University of Chicago Press.

Ribar, David C. and Mark Ottoni-Wilhelm. 2002. "Altruistic and Joy-of-Giving Motivations in Charitable Behavior." *Journal of Political Economy* 110 (2): 425-457.

Roberts, Russell D. 1984. "A Positive Model of Private Charity and Public Transfers." *Journal of Political Economy* 92 (1): 136-148.

Turner, Robert W. and Reed Walker. (2006). "Explaining the Determinants of Operating Budgets at U.S. National Parks." *Public Finance Review* 34 (5): 551-573.

United States Department of the Interior (US DOI). 2019. "Budget Justifications and Performance Information for Fiscal Year 2018: National Park Service." https://www.nps.gov/aboutus/upload/fy-2018-nps-greenbook.pdf/ (accessed February 26, 2019).

Warr, Peter G. 1982. "Pareto Optimal Redistribution and Private Charity." *Journal of Public Economics* 19: 131-138.

Figures and Tables

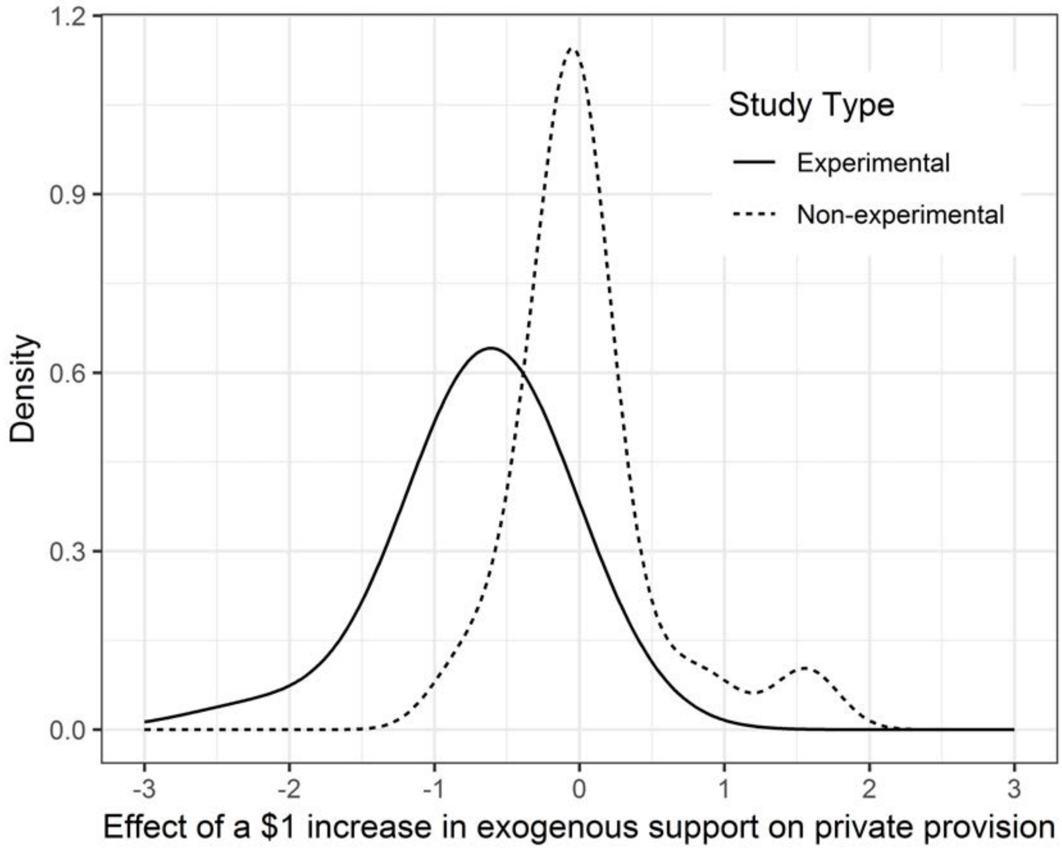

Figure 1: Kernel density function of crowding effect estimates (negative is crowding out, positive is crowding in) from the literature, shown separately for experimental and non-experimental studies, and based on 325 estimates from 54 studies collected by de Wit and Bekkers (2017).

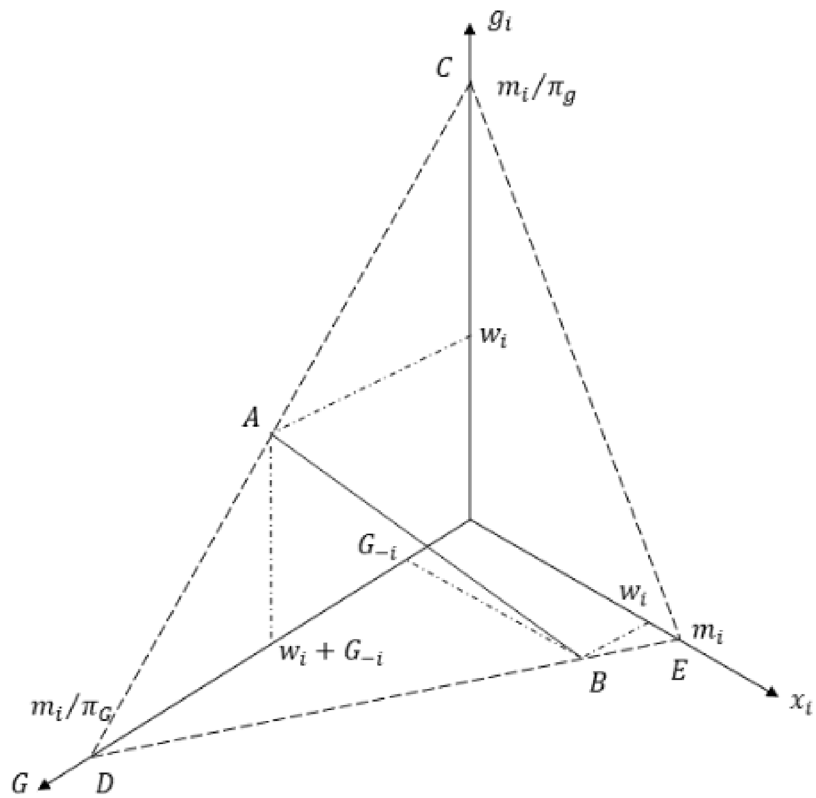

Figure 2: The budget frontier, along with virtual magnitudes, for impure altruism in three-dimensional utility space.

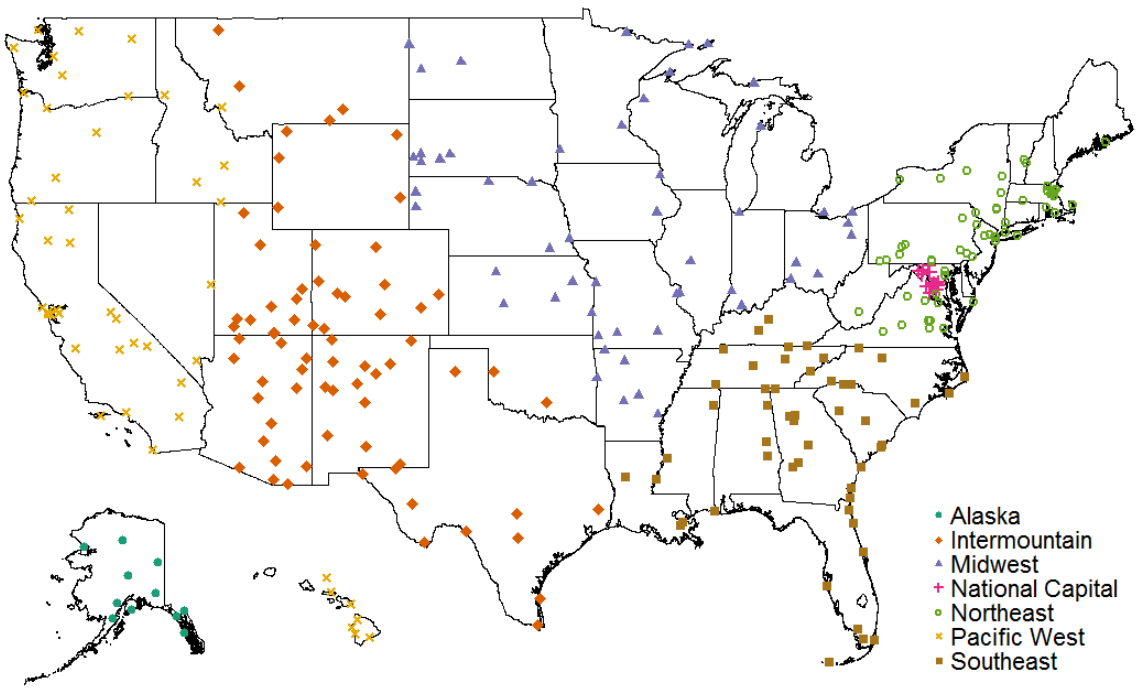

Figure 3: Geographic distribution of the 326 parks used in the analysis and managed by the National Park Service, by administrative region, excluding 5 parks in American Samoa, the Mariana Islands, and Puerto Rico.

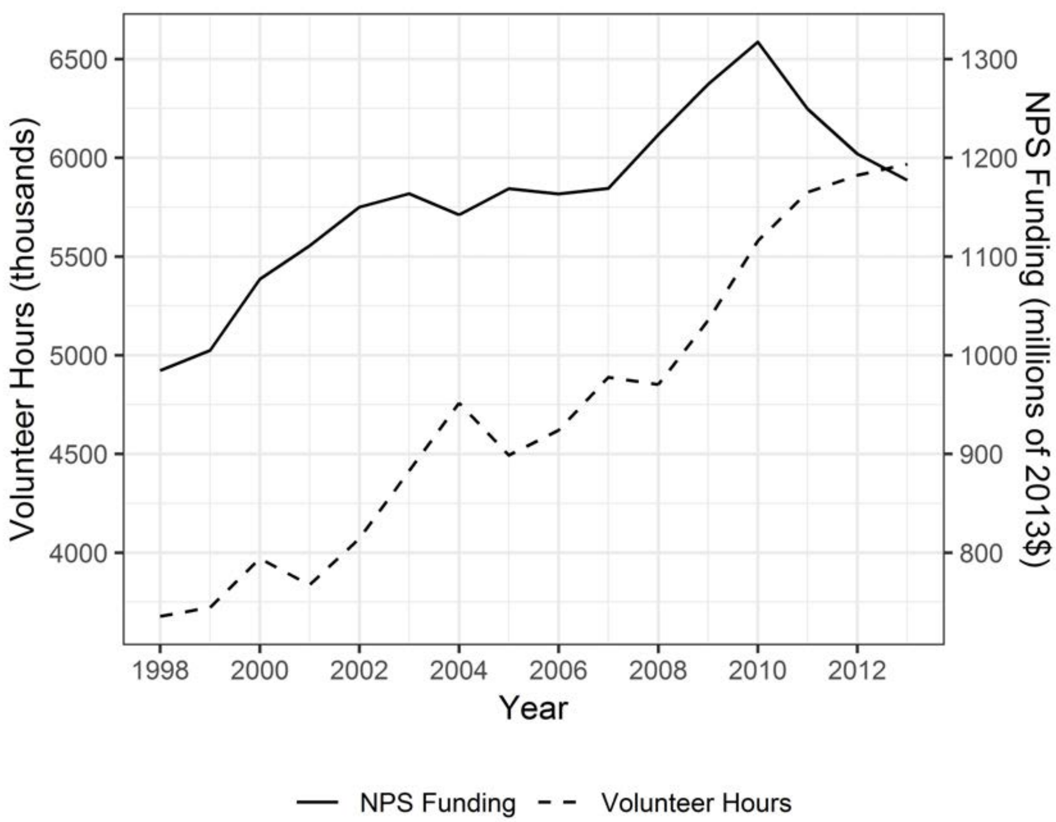

Figure 4: Trend in annual National Park Service volunteer hours and funding to parks

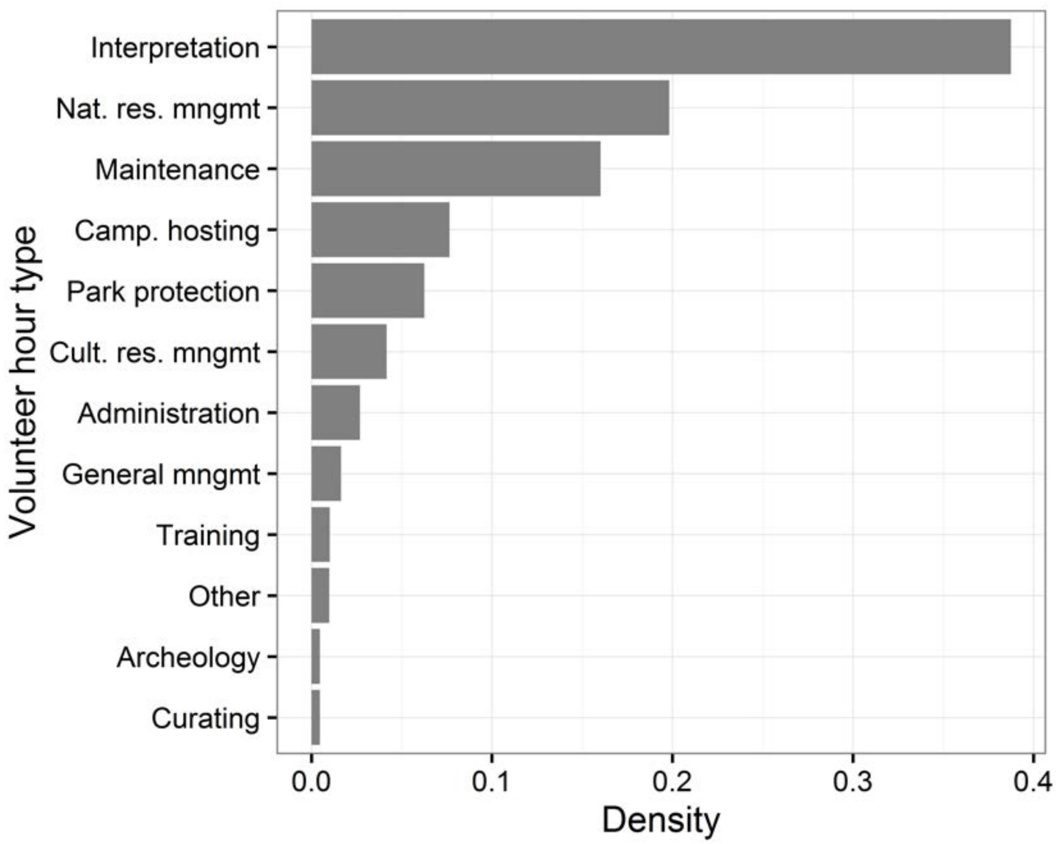

Figure 5: Breakdown of volunteer hours by type for all parks, 1998-2013

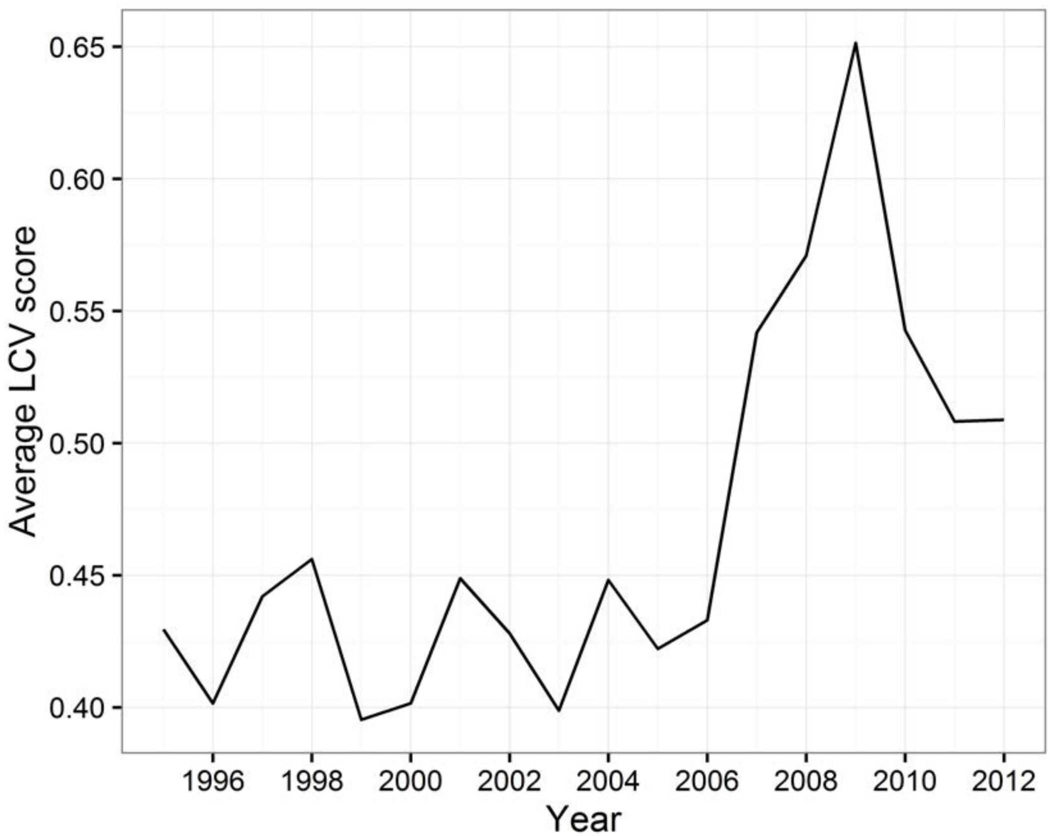

Figure 6: Time series of the average annual LCV score based on all members of the U.S. House and Senate Appropriations Committees

Table 1: Summary of the qualitative sign of comparative static results compared with standard assumptions for impure altruism

Comparative Statics	General		Standard Assumptions
	Substitutes	Complements	
$\frac{\partial G^*}{\partial w_i}$	> 0	?	$0 < f_{i1} < 1$
$\frac{\partial G^*}{\partial G_{-i}}$	> 0	?	$0 < f_{i1} + f_{i2} < 1$
$\frac{\partial G^*}{\partial G_{-i}} - \frac{\partial G^*}{\partial w_i}$	?	?	$f_{i2} > 0$
Crowding effects			
$\frac{\partial g_i^*}{\partial G_{-i}}$	$?(> -1)$	?	$-1 < f_{i1} + f_{i2} - 1 < 0$
$\frac{\partial g_i^*}{\partial G_{-i}}\Big\rvert_{dG_{-i}=-dw_i}$	< 0	?	$-1 < f_{i2} - 1 < 0$

Notes: Substitutes and complements refer to cases where G and g_i are net (Hicksian) substitutes or complements in an unrestricted utility maximization problem. Standard assumptions refers to the conditions assumed in the existing literature on the theory of impure altruism.

Table 2: Summary statistics

	All Parks	Environmental Parks	
		Yes	No
$Hours_{it}$: Volunteer Hours per Year	15,686	29,606	8,454
	(30,236)	(46,110)	(11,497)
Value of Annual Volunteer Hours ($ 1,000s)	346	656	186
	(674)	(1,029)	(256)
$Budget_{it}$: Annual Budget ($ 1,000s)	3,844	6,395	2,520
	(4,999)	(6,513)	(3,288)
$Visits_{it}$: Annual Visits (1,000s)	822	1,355	547
	(1,776)	(2,116)	(1,501)
FTE_{it}: Annual Paid Staff FTE (per 1,000 visits)	0.334	0.280	0.361
	(1.734)	(0.757)	(2.067)
Number of parks	326	105	221
Observations	4,808	1,644	3,164

Notes: Standard deviations are reported in parentheses. Parks are classified as environmental according to our primary definition of whether volunteerism in the category of natural resource management is strictly positive for every year. The value of volunteer hours is calculated using the annual value of volunteer time from the Bureau of Labor Statistics (see text). The value of volunteer hours and the annual budget are reported in 2013 dollars.

Table 3: First stage results for IV estimation

	(1)	(2)	(3)
Z_{it}	0.215***	0.187***	-0.034
	(0.059)	(0.071)	(0.034)
$Z_{it} \times 1[Envr]_i$		0.034	0.288***
		(0.078)	(0.054)
$Visits_{it}$	-0.038	-0.047	-0.029
	(0.264)	(0.258)	(0.167)
FTE_{it}	2.075	2.061	0.419
	(2.412)	(2.402)	(0.913)
Park FE	✓	✓	✓
Region-Year FE	✓	✓	✓
F-stat on excluded instruments	13.25	6.93	15.26
Observations	4,808	4,808	4,808

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Notes: Column 1 is the first stage regression of specification (14); the dependent variable is annual park funding in thousands of 2013 dollars. Columns 2 and 3 are the first stage regressions of specification (15); the dependent variable is, respectively, annual park budget in thousands of 2013 dollars, and the interaction of budget with the indicator variable for an environmental park. Z_{it} is the instrumental variable described in equation (16). Robust standard errors clustered by park are reported in parentheses, and we report cluster-robust F-statistics on the excluded instruments. The Kleibergen-Paap rk Wald F-statistic is 6.4 for the joint test in columns 2 and 3.

Table 4: Overall estimates of the average crowding effect

	OLS			IV		
	(1)	(2)	(3)	(4)	(5)	(6)
$Budget_{it}$	2.509*	2.511*	2.445**	12.731***	12.410**	9.092**
	(1.500)	(1.494)	(1.152)	(4.779)	(4.826)	(3.566)
$Visits_{it}$		0.231	2.220		1.024	1.101
		(2.202)	(1.433)		(3.552)	(1.782)
FTE_{it}		-50.055**	-53.943**		-77.237**	-56.159**
		(24.416)	(22.487)		(33.669)	(22.134)
Park FE	✓	✓	✓	✓	✓	✓
Region-Year FE	✓	✓	✓	✓	✓	✓
Observations	4,808	4,808	3,868	4,808	4,808	3,868

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Notes: The dependent variable is total volunteer hours by park-year. Columns 1-3 are estimated with OLS, and columns 4-6 are estimated using the IV approach described in the main text. Columns 1, 2, 4, and 5 include the full sample; columns 3 and 6 are estimated on the 1998-2010 time period, excluding the post-recession years. Robust standard errors clustered by park are reported in parentheses.

Table 5: Heterogeneous crowding effects by environmental and non-environmental parks

	OLS			IV		
	(1)	(2)	(3)	(4)	(5)	(6)
$Budget_{it}$	-0.177	-0.177	0.732	3.511	1.348	0.684
	(0.437)	(0.440)	(0.554)	(4.001)	(3.666)	(2.976)
$Budget_{it} \times 1[Envr]_i$	4.435***	4.435***	2.321*	9.732**	11.554***	8.816**
	(1.560)	(1.572)	(1.362)	(4.116)	(4.417)	(3.731)
$Visits_{it}$		-0.025	2.321		0.032	1.793
		(1.977)	(1.445)		(2.306)	(1.699)
FTE_{it}		-47.915**	-56.772**		-60.520***	-62.470***
		(23.534)	(23.245)		(22.002)	(23.720)
Park FE	✓	✓	✓	✓	✓	✓
Region-Year FE	✓	✓	✓	✓	✓	✓
Observations	4,808	4,808	3,868	4,808	4,808	3,868

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Notes: The dependent variable is total volunteer hours by park-year. Columns 1-3 are estimated with OLS, and columns 4-6 are estimated using the IV approach described in the main text. Columns 1, 2, 4, and 5 include the full sample; columns 3 and 6 are estimated on the 1998-2010 time period, excluding the post-recession years. $1[Envr]_i$ is an indicator variable for an environmental park. Robust standard errors clustered by park are reported in parentheses.

Table 6: Crowding effects separately for outside and inside volunteer hours

	OLS		IV	
	(1)	(2)	(3)	(4)
Panel A: Outside Hours				
$Budget_{it}$	1.742	1.740	9.235**	9.270**
	(1.161)	(1.151)	(3.980)	(4.054)
$Visits_{it}$		-0.300		0.304
		(1.880)		(2.723)
FTE_{it}		-49.856**		-70.534**
		(19.178)		(28.230)
Panel B: Inside Hours				
$Budget_{it}$	0.766	0.771	3.066*	3.140*
	(0.748)	(0.744)	(1.783)	(1.810)
$Visits_{it}$		0.531		0.721
		(1.439)		(1.653)
FTE_{it}		-0.199		-6.703
		(9.647)		(9.972)
Park FE	✓	✓	✓	✓
Region-Year FE	✓	✓	✓	✓
Observations	4,808	4,808	4,808	4,808

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Notes: The dependent variable is volunteer hours devoted to outside tasks in Panel A and inside tasks in Panel B. Outside hours include archeology, campground hosting, interpretation, resource management, and park protection. Inside hours include general management, curating, administration, maintenance, training, and other. Columns 1-2 are estimated with OLS, and columns 3-4 are estimated using the IV approach described in the main text. Robust standard errors clustered by park are reported in parentheses.

Table 7: Heterogeneous crowding effects by park and hour types

	OLS		IV	
	(1)	(2)	(3)	(4)
Panel A: Outside Hours				
$Budget_{it}$	-0.153	-0.163	2.111	1.935
	(0.365)	(0.376)	(3.282)	(3.503)
$Budget_{it} \times 1[Envr]_i$	3.129**	3.139**	7.505**	7.662**
	(1.431)	(1.450)	(3.737)	(3.580)
$Visits_{it}$		-0.481		-0.354
		(1.825)		(2.026)
FTE_{it}		-48.341***		-59.447***
		(18.187)		(19.447)
Panel B: Inside Hours				
$Budget_{it}$	-0.024	-0.014	-0.788	-0.586
	(0.178)	(0.185)	(1.335)	(1.388)
$Budget_{it} \times 1[Envr]_i$	1.306	1.296	4.064**	3.892**
	(0.994)	(0.977)	(1.790)	(1.799)
$Visits_{it}$		0.455		0.387
		(1.371)		(1.400)
FTE_{it}		0.426		-1.072
		(9.747)		(9.432)
Park FE	✓	✓	✓	✓
Region-Year FE	✓	✓	✓	✓
Observations	4,808	4,808	4,808	4,808

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Notes: The dependent variable is volunteer hours devoted to outside tasks in Panel A and inside tasks in Panel B. Outside hours include archeology, campground hosting, interpretation, resource management, and park protection. Inside hours include general management, curating, administration, maintenance, training, and other. Columns 1-2 are estimated with OLS, and columns 3-4 are estimated using the IV approach described in the main text. $1[Envr]_i$ is an indicator variable for an environmental park. Robust standard errors clustered by park are reported in parentheses.

Appendix Tables

Table A.1: Summary statistics with the alternative definition of an environmental park

	All Parks	Environmental Parks	
		Yes	No
$Hours_{it}$: Volunteer Hours per Year	15,686	18,187	8,669
	(30,236)	(34,062)	(12,546)
Value of Annual Volunteer Hours ($ 1,000s)	346	400	190
	(674)	(752)	(277)
$Budget_{it}$: Annual Budget ($ 1,000s)	3,844	4,223	2,783
	(4,999)	(5,158)	(4,352)
$Visits_{it}$: Annual Visits (1,000s)	822	890	632
	(1,776)	(1,910)	(1,314)
FTE_{it}: Annual Paid Staff FTE (per 1,000 visits)	0.334	0.314	0.388
	(1.734)	(1.835)	(1.415)
Number of parks	326	233	93
Observations	4,808	3,545	1,263

Notes: Standard deviations are reported in parentheses. Parks are classified as environmental according to our secondary definition of whether a park's official mandate includes activities related to environmental conservation. The value of volunteer hours is calculated using the annual value of volunteer time from the Bureau of Labor Statistics (see text). The value of volunteer hours and the annual budget are reported in 2013 dollars.

Table A.2: Alternative specifications to test for robustness

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
$Budget_{i,t-1}$	13.860**	3.548					
	(5.411)	(3.971)					
$Budget_{i,t-1} \times 1[Envr]_i$		10.604**					
		(4.524)					
$Budget_{it}$			11.629**	-1.217	10.513**	-0.799	1.538
			(4.903)	(3.957)	(4.712)	(3.769)	(3.745)
$Budget_{it} \times 1[Envr]_i$				13.226***		12.157***	12.506**
				(4.642)		(4.113)	(5.590)
$Visits_{it}$	-1.894	-1.100	2.218	0.108	1.148	0.083	-0.320
	(2.641)	(2.653)	(3.601)	(1.722)	(3.166)	(2.069)	(2.473)
FTE_{it}	-53.714**	-56.936***	-304.209	-3.628	-84.550**	-65.702***	-69.362***
	(23.866)	(20.236)	(245.550)	(230.726)	(33.038)	(21.881)	(25.854)
Park FE	✓	✓	✓	✓	✓	✓	✓
Region-Year FE	✓	✓	✓	✓	✓	✓	✓
Lagged Budget	✓	✓					
Include Lagged Visits			✓	✓			
Include Time Trends for Large/Small Parks					✓	✓	
More General Definition of Envr. Parks							✓
K-P F-stat	15.32	6.73	14.03	7.44	13.34	7.36	6.47
Observations	4,529	4,529	4,109	4,109	4,808	4,808	4,808

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Notes: The dependent variable is total volunteer hours by park-year. All specifications are estimated using the IV approach described in the main text. Models (1) and (2) include the park budgets lagged one year. Models (3) and (4) include two years of lagged visitation. Models (5) and (6) include separate linear time trends for parks with above and below median 1995 budgets. Model (7) estimates the heterogeneous park effects with our secondary definition of an environmental park based on its stated mandate. Robust standard errors clustered by park are reported in parentheses. We report the cluster-robust Kleibergen-Paap rk Wald F-statistic in the first stage.

Table A.3: Alternative instruments

	(1)	(2)	(3)	(4)	(5)	(6)
$Budget_{it}$	11.119**	-4.685	12.516***	3.483	8.059***	1.796
	(4.684)	(4.545)	(4.025)	(3.691)	(1.945)	(2.435)
$Budget_{it} \times 1[Envr]_i$		16.140***		9.708**		6.858***
		(4.666)		(3.876)		(2.529)
$Visits_{it}$	0.921	-0.493	1.033	0.220	0.675	0.111
	(3.327)	(2.182)	(3.582)	(2.456)	(2.778)	(2.112)
FTE_{it}	-73.691**	-49.373**	-77.527**	-64.199**	-65.288**	-56.197**
	(31.559)	(20.469)	(34.200)	(25.237)	(29.612)	(24.558)
Park FE	✓	✓	✓	✓	✓	✓
Region-Year FE	✓	✓	✓	✓	✓	✓
All Congressional LCV Scores IV	✓	✓				
Fish and Wildlife Service IV			✓	✓		
NPS Budget For Other Regions IV					✓	✓
K-P F-stat	14.04	4.98	10.07	1.81	16.05	1.49
Observations	4,808	4,808	4,808	4,808	4,808	4,808

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Notes: The dependent variable is total volunteer hours by park-year. All specifications are estimated using the IV approach described in the main text. Models (1) and (2) are estimated using the instrumental variable based on the LCV scores for all U.S. Congressional members, excluding those representing each park. Models (3) and (4) are estimated using the instrumental variable based on the budget for the U.S. Fish and Wildlife Service. Models (5) and (6) are estimated using the instrumental variable based on the total NPS budget, exclusive of the region where each park is located. Robust standard errors clustered by park are reported in parentheses. We report the cluster-robust Kleibergen-Paap rk Wald F-statistic in the first stage.