Peer Effects in Water Conservation: Evidence from Consumer Migration*

Bryan Bollinger
Duke University

Jesse Burkhardt
Colorado State University

Kenneth Gillingham
Yale University

December 8, 2017

Abstract

Social interactions are widely understood to influence consumer decisions in many choice settings. This paper identifies causal peer effects in water conservation utilizing variation from consumer migration. Using satellite data and a machine learning approach, we show that the water conservation effects can be attributed to peer effects in the diffusion of dry landscaping. We examine households that receive free irrigation water from landscaping to provide evidence that without a price signal, peer effects are muted. These results inform water use policy in many areas of the world threatened by recurring drought conditions.

Keywords: social interactions; diffusion; water policy.
JEL classification codes: Q42, Q48, L13, L25, O33, O25.

*Bollinger: Fuqua School of Business, Duke University, 100 Fuqua Drive, Durham, NC 27708, phone: 919-660-7766, e-mail: bryan.bollinger@duke.edu. Burkhardt: Department of Agricultural and Resource Economics, Colorado State University, 1200 Center Ave. Mall, Fort Collins, CO 80523, phone: 503-312-8943, email: jesse.burkhardt@colostate.edu. Gillingham: School of Forestry & Environmental Studies, Department of Economics, School of Management, Yale University, 195 Prospect Street, New Haven, CT 06511, phone: 203-436-5465, e-mail: kenneth.gillingham@yale.edu. The authors are grateful for conversations and comments from Kerry Smith and Michael Hanemann, and from seminar participants at Arizona State, Northwestern, and UCLA. The authors would also especially like to thank the staff at the Phoenix Metropolitan Water District, including Doug Frost, Darren Sversvold, Soe Myint, and Jamie Campbell.

1 Introduction

Over the 20th century, global water use seen a sixfold increase and by 2025, the U.N. Environmental Program estimates that over two-thirds of the world's population will live in water-stressed regions. Concerns about the water availability are equally important in the many regions of the United States. The California droughts of 2012 and 2015 led to combined economic losses of approximately $5.2 billion, and only a few years earlier droughts in the Southwest and Midwest led to losses of $20 billion.[1]. The U.S. Environmental Protection Agency predicts up to a 40% decrease in snow runoff and soil moisture in parts of the Western United States by 2050.[2] Residential water demand is also of further interest due to rapid change in both technologies and practices. For example, between 1990 and 2010, the diffusion of higher water efficiency appliances and a shift towards dry landscaping instead of grass has reduced the per capita residential water use by roughly 10%.[3]. This decline in per capita water usage is a positive development for concerns about water availability, but can pose challenges for municipalities that designed water systems for greater demand. Thus, understanding residential water demand and the diffusion of technologies and practices that influence demand is of major interest to policymakers.

Social interactions have been shown to play a pivotal role in the diffusion of many new technologies and practices and undergird classic economic models of technology diffusion (Griliches 1957; Bass 1969; Rogers 1995). The idea that individuals learn from their peers, neighbors, or friends to adopt behaviors or technologies has been explored in settings ranging from agriculture (Foster and Rosenzweig 1995; Conley and Udry 2010) to foreclosures (Towe and Lawley 2013) to schooling (Sacerdote 2001; Graham 2008). In the environmental realm, such 'peer effects' have been shown in the adoption of solar photovoltaic panels (Bollinger and Gillingham 2012; Graziano and Gillingham 2015) and hybrid vehicles (Narayanan and Nair 2013; Heutel and Muehlegger 2015).

[1] See https://www.ama.org/publications/eNewsletters/Marketing-News-Weekly/Pages/economic-loss-us-droughts.aspx and http://www.cnbc.com/2015/03/03/california-drought-seen-having-worsening-3-billion-economic-impact-in-2015.html

[2] See https://19january2017snapshot.epa.gov/climate-impacts/climate-impacts-southwest_.html

[3] See https://water.usgs.gov/watuse/data/2010/index.html

This paper is the first to identify causal peer effects in water consumption, with a focus on landscape choice. Our identification strategy relies on consumer migration that leads new households to move into the peer group of each household. We use water billing data from over 300,000 households in Phoenix–one of the hottest and driest cities in the United States–along with housing transaction data to find that a 10% reduction in peer water consumption leads to a 1.65% or 207 gallon per month reduction in a household's water consumption. To understand the mechanism underlying this result, we use a machine learning approach to code the greenness of a household's landscaping using satellite data. We show that the peer effects in water consumption are driven almost entirely from changes in landscaping. Further, we use the Salt River Project, which provides extremely-heavily discounted irrigation water for most of a household's outdoor landscaping needs, to provide evidence indicating that a positive price is necessary for these peer effects to influence consumer behavior. We also explore heterogeneity in these effects to provide guidance for targeted policy.

It is well understood that a causal peer effects are challenging to identify. These challenges include endogenous group formation leading to self-selection of peers, correlated unobservables, and simultaneity issues (Manski 1993; Brock and Durlaf 2001; Moffitt 2001; Hartmann et al. 2008). This study addresses endogenous group formation using a rich set of geographic and temporal fixed effects. We address the other concerns using an instrumental variables strategy based on exogenous shocks to the landscape from movers to a neighborhood. New residents to a particular house often change the landscaping of that property within a short time after the move. At the same time, there is a broader trend towards dry landscaping occurring in Phoenix, as we will document. Thus, we observe a much higher probability of a switch from green to dry landscaping just after a housing transaction. Our identification strategy relies on these shocks to landscaping from consumer migration to instrument for the average water consumption and landscape decisions of a household's peer group, after already focusing on the variation remaining after inclusion of time-varying geographic fixed effects.

This work contributes to several literatures. Most directly, it adds new well-identified evidence to the large and growing literature on peer effects in the diffusion of consumer behaviors. But it also adds to the literature on how information transmission, in our case through social learning, can influence consumer decisions about energy and water use. For example, Bollinger (2015) shows that information and peer demonstrations of an environmentally-friendly technology can have a large effect on individual adoption of green technology. Several papers explore how social norm messages aimed at energy conservation can reduce energy use (e.g., Allcott 2011; Allcott and Rogers 2014; Ayres et al. 2012; Costa and Kahn 2013; Dolan and Metcalfe 2015; Tsvetanov and Gillingham 2017) or how prosocial appeals influence energy use relative to economic incentives (e.g., Reiss and White 2008; Ito et al. 2017; Burkhardt et al. 2017). Our paper is novel in being the first to explore how consumer behavior change due to social interactions is influenced by economic incentives. Jessoe and Rapson (2014) show that households are three standard deviations more responsive to temporary price increases when provided with high frequency information on electricity usage and Dolan and Metcalfe (2015) indicate that the effect of financial incentives disappears when information on social norms is provided. In the context of residential water demand, Ferraro and Price (2013) show that social comparison messages are the most effective among the least price sensitive households. Our results approach this in a different way in showing that information transmission from social interactions is vastly muted without a price signal to make the information transmission worth it.

Our findings contribute to active policy discussions. Water districts must make large upfront investments in infrastructure and face continual challenges in planning. Understanding the speed and pattern of the diffusion of water conservation activities, such as a transition from green to dry landscaping, is immediately valuable for water planning purposes. Neighborhoods with a small number of transitions in a given year may be expected to observe an greater number transitions in the future relative to other neighborhoods. Recognizing this fact can lead to targeted policies. For example, one-time

targeted information campaigns or subsidies for dry landscaping may be useful in fast-growing areas that are straining the water system, while such information may be best withheld from areas that have too much capacity. Further, even more refined targeted campaigns may be possible. We explore the heterogeneity to show that pet owners show minimal peer effects in landscaping, consistent with a high preference for green space, while household close to a park or school display greater peer effects in landscaping, consistent with a lower preference for green space due to greater access to nearby public green space.

The remainder of the paper is organized as follows. In the next section we provide some background on water use in Arizona and describes our unique dataset, which combines remote sensing images, water bills, and housing transactions. Section 3 presents the model and discusses our identification strategy. Section 4 describes the results and mechanisms underlying the peer effect. Section 5 concludes.

2 Background and Data

2.1 Background on Water Use in Arizona

Phoenix receives only 8 inches of rain a year on average. Residential water in Phoenix comes from three sources: wells owned by individuals or the city, the Salt and Verde rivers, and the Colorado river.[4] These sources are both constrained and politically contentious. Moreover, between 1980 and 2010, the population of Phoenix increased by approximately 83%, leading to an increase in residential water use of 23%.[5]

Fortunately for Phoenix and many arid regions of the United States, water use per household has in general been trending downwards over time. Figure 1 shows the average water use per household in Phoenix between 1986 and 2012. While there was a slight increase in the late 1990's, there is a clear downward trend throughout the entire time

[4]For details see http://archive.azcentral.com/members/Blog/ShaunMcKinnon/123051
[5]See https://www.biggestuscities.com/city/phoenix-arizona

period. The broader decline in per household water use is usually primarily attributed to improved water efficiency of appliances, such as low-flow toilets and shower heads or water-efficient washers/driers and dishwashers. Yet there is a growing consensus among practitioners that the low hanging fruit in improvements in water efficiency has mostly been picked.

More recently, water districts have thus been focusing on behavior change and outdoor water use. Water conservation campaigns have become popular (Ferraro and Price 2013) and some areas have changed water pricing. Promoting a shift to dry landscaping, with less grass and more drought-tolerant plants, is another option that has been adopted in numerous arid water districts. Further, in some cases this transition to dry landscaping is occurring without policy action. As we will see shortly, Phoenix has seen just such a trend in recent years.

2.2 Data

The foundation of our data set is water billing data from the City of Phoenix water department for all single-family households served by the water department between 2004 and 2012 (316,714 households). These contain the address of the land parcel, allowing us to geocode all parcels. They also contain the total annual water consumption for each household, with the exception of parcels that are eligible to receive a defined amount of highly reduced-cost flood irrigation water from the Salt River Project (SRP), a relic from previous agricultural uses. This flood irrigation water is only suitable for outdoor uses, but not all eligible houses use the SRP water and many eligible houses supplement the SRP water with municipal water.[6] We do not observe the use of SRP irrigation water, but can note that SRP water provision did not change during our empirical setting. Roughly

[6]For a household to use the SRP water, it must construct berms to direct the water and the water is not suitable for drinking. The SRP water is provided to households based on plot size and is distributed in 45 minute increments. Households only pay a small fee for the water delivery; currently, this fee is on the order of $5 per month for most households and has not appreciably changed in recent years. We find that annually, off-project houses use approximately 12 cubic feet of water more than on project houses on average. However, there is no statistically significant correlation between annual municipal water consumption and SRP eligibility when lot size is controlled for.

43% of households in Maricopa County are within the boundaries of the SRP.

We complement the water billing data with satellite images provided by the City of Phoenix in order to develop a measure of landscape greenness. These images, taken once a year in the fall have resolutions ranging from 0.32 to 0.8 feet. All years are available except 2010, which is only available at a lower resolution. Each image has three color bands (red, green, and blue). There are standard software packages for developing vegetation indices, such as the Normalized Difference Vegetation Index, but our satellite images lack the infrared band needed for such standard packages. Therefore, we used a machine learning approach to train the computer to find green pixels through a series of iterations. The approach we used is a supervised maximum likelihood classification approach developed by the company Imagine Software, Inc. This approach achieved an accuracy of approximately 85-90% in our comparisons of the machine learning approach to hand-coded green pixels.[7] Figure 3 provides an example of the output of the classification process. The photo on the left shows several randomly chosen parcels while the photo on the right shows the same parcels with the pixels the computer designates as green landscaping, including tree tops, grasses, and other vegetation, highlighted in green. Figure 2 shows a map of Phoenix with the water department territory, Salt River project territory, and the remote sensing subsample of 67,370 parcels. In the Appendix, we demonstrate that these parcels are quite representative of the full sample.

As is common in remote sensing work, further data cleaning is necessary to compare satellite images taken at different times. For instance, images may contain more or less shadow and more or less haze depending exactly on when they were taken. The cameras or camera settings used may have also been slightly different in each photo. Thus, for consistency across years, we systematically adjust the remote sensing data by focusing on the variation in the landscape greenness measure that can be explained by the water consumption data. Specifically, we regress landscape greenness on water consumption and use the fitted values of landscape greenness, thus allowing us to focus on the relevant variation in landscape greenness. This data adjustment is discussed in more detail in

[7] We hand-coded over 100 parcels for the comparison.

Appendix A. Figure 4 plots the average percentage of green landscapes in our adjusted data over time, not surprisingly indicating the same clear downward trend as in the water consumption data.

Our next data source is the Maricopa County Assessors Office, which provided data on housing sales and other other physical housing characteristics including pool size, lot size, construction date, home size, and garage size. Importantly, we observe the date at which the housing transactions occur and the address of the parcel, allowing us to match these data with the previous datasets.

For both the water billing data and the landscape data, we geocode all addresses. To create a measure of peer decisions, we create a radius around the center of each household's parcel. In our primary specifications, we use a radius of 500 feet, but we also explore different radii in our robustness checks to show that the effect is relatively constant across a range of different radii. We then create variables for the average water use and average landscape greenness for all surrounding households within the 500-foot radius. Using the moving data we also create a variable for the fraction of households sold within that radius and during that year. Table 1 presents the summary statistics for both the water consumption data, landscape data, and moving data. It shows that just over 4% of the water consumption sample moves each year and that the mean fraction of landscape greenness is 37% (with the mean taken over household parcels). All variables show substantial variation.

Our final data source is the marketing company Acxiom. Acxiom provides household-level demographic information collected from a variety of sources, including public records and proprietary suppliers. These include data on physical household characteristics, such as the home size, as well as variables such as income, age, political party affiliation, and indices of interests based on magazine subscriptions. These data are useful for exploring the heterogeneity in our estimates. In the appendix, we present summary statistics for both the household characteristics from the Assessors Office and the Acxiom demographic variables.

3 Empirical Specification

3.1 Challenges in Identifying Peer Effects

Peer effects are notoriously challenging to identify. A first question in any model of social interactions is how to define the peer group. By defining the peer group membership too broadly could pick up sufficient heterogeneity in the group that it leads to a spurious correlation. Indeed, some studies are even based on careful definition of the peer group (e.g., Bertrand et al. 2000). In our setting, we are interested in how peer effects influence the choice of landscaping and water usage. While water usage may not be visible, landscaping is usually highly visible. This lends itself to a definition of peer group based on spatial proximity to the focal household parcel. In a similar setting in Towe and Lawley (2013), neighbors are defined as the nearest 13 and nearest 25 neighbors by distance. Because there is variation in parcel sizes, we prefer a measure based on the radius around the focal household parcel (our 500' radius include 23 neighbors on average). This choice of geographic peer group is also common in the literature (Topa 2001; Arzaghi and Henderson 2007; Bell and Song 2007; Manchanda et al. 2008; Choi et al. 2010; McShane et al. 2012; Nam et al. 2010; Narayanan and Nair 2013). Of course, households can be expected to have other social groups as well, such as those relating to family, friends, schools, and jobs. So we view our measure as a minimal measure of the social group relevant to water and landscape decisions.

Once the peer group has been defined, identification of peer effects also relies on addressing the further classic concerns: self-selection of peers, correlated unobservables, and simultaneity issues. The challenge of self-selection of peers (sometimes described as 'homophily') is a major concern for identifying the relationship between an individual's decision and the average decisions by the peer group because individuals endogenously choose their social networks. In the case of housing choice, consumers with similar preferences can be expected to sort into neighborhoods. Empirical studies have have often addressed this issue with a rich set of fixed effects. Correlated unobservables refer to the

many other factors that may influence both the focal consumer and peers. For example, if there is an economic downturn facing all households in a neighborhood, their decisions may all appear to be aligned, but this alignment is due to the conditions faced by the households, rather than peer effects. Studies address this issue in various ways, including time-varying fixed effects or another source of exogenous variation that shifts the peer characteristics but does not shift the individual characteristics–an instrument.

Simultaneity issues (sometimes called 'reflection') refers to the concern that just as peers may influence the focal household, the focal household influences the peers. At the extreme this can lead to mathematical non-identification. It can be thought of as an endogeneity issue analogous to the classic simultaneity of supply and demand. In the supply/demand context, the standard instrumental variables approach requires an exclusion restriction, such as a supply shifter to identify demand or a demand shifter to identify supply. Empirical studies often use instruments to address this issue or alternatively use lagged values of the peer group variable.

For both the issue of simultaneity and correlated unobservables, the fundamental challenge is one of research design. Angrist (2014) clearly points out the ideal characteristics of a well-identified peer effects study: "Research designs that manipulate peer characteristics in a manner unrelated to individual characteristics provide the most compelling evidence on the nature of social spillovers." Our research design overcomes the above challenges by both focusing on variation within households and within zip code-by-year and by exploiting an exogenous shock that fulfills the criteria laid out by Angrist (2014).

3.2 Our Empirical Strategy

Consider a classic linear-in-means specification that models the water consumption or landscape choice of household i in year t as a function of the peer group's aggregate choices, whether the household is moving in that year, time-invariant household charac-

teristics, and time-varying characteristics about the local neighborhood:

$$y_{it} = \theta \bar{y}_{it} + \delta m_{it} + \eta_i + \phi_{tz} + \epsilon_{it}, \tag{1}$$

where y_{it} is the household's choice (i.e., either water consumption or landscape choice). If we denote household i's peer group as P_i, then $\bar{y}_{it} = \frac{1}{|P_i|} \sum_{i' \in P_i} y_{it}$ is the average of the choices of the household's peers.[8] m_{it} is a dummy variable for whether the household just moved to the parcel they are occupying in that year (i.e., a housing transaction), for we would expect landscape changes when people first move into a house due to differences in preferences. η_i contains time invariant household characteristics, which we model as a household fixed effect (i.e., a fixed effect for each parcel x owner combination, so that there is a different fixed effect after a sale). ϕ_{tz} captures time-varying factors such as localized economic shocks or major new development in a neighborhood, and we model this with zip code x year fixed effects.

One clear implication of this model is that as long as δ, the coefficient on m_{it}, is not zero, we would expect the intensity of migration into a neighborhood to also influence $\bar{y}_{it}$. At the same time, the intensity of migration into a neighborhood acts as a shifter of the peer group's characteristics in a manner unrelated to the individual characteristics because it only affects the peer group. This suggests the use of the average migration into the peer group, $\bar{m}_{it}$, as an instrument for $\bar{y}_{it}$ in (1).

The key identification assumption in this research design is that there are no unobserved factors that influence both the decision of potential neighbors to relocate into that neighborhood and the decision of the individual household to change their water consumption or landscape. Our time-varying localized fixed effects are crucially important for allowing us to focus on the idiosyncratic shocks (off the mean) to water consumption or landscapes after conditioning on broader neighborhood changes. The remaining variation after including time-varying localized fixed effects is due to individuals deciding to change their water consumption or landscapes after individual shocks, such as a decision

[8] In the incomplete information framework of Manski (1993), $\bar{y}_{it}$ would be represented by $\mathbb{E}[y_{gt}]$, where g refers to the group.

to renovate, individual-specific promotion or loss of a job, etc.

This identification strategy overcomes the three primary challenges discussed above. Household fixed effects immediately address self-selection of peers based on household-specific preferences. Both correlated unobservables and simultaneity are endogeneity concerns, and our instrumental variables strategy addresses these concerns after focusing on the remaining variation after removing the time-varying localized fixed effects. Of course, for our research strategy to work, the instrument cannot be weak.

As would be expected, we find a strong relationship between the fraction of movers into the peer group and the peer group's decision. Table A.6 in Appendix D includes our first stage regression results, showing that there is a strong relationship between the instrument and the peer group's average water consumption and landscape decisions. The F-statistic for the first stage is well above 20, and the instruments consistently pass under-identification and weak identification tests at the 1% level. The economic explanation for this relationship is that when people move into a new house, they are likely to use the opportunity to change things about the house, including the landscaping. We find a negative relationship between the fraction of sold houses in the peer group and water consumption and landscape decisions, suggesting that households that move transition from greener landscapes to drier landscapes. This would be consistent with the broader trend towards drier landscapes in the Southwest, and consumer migration provides the exogenous shock to the transitions.

4 Results

This section is divided into four subsections. First, we present the results of our primary peer effects specifications for total water consumption and landscape choices. Second, we provide evidence that the peer effect is influenced by economic incentives by separately estimating the peer effect for on and off Salt River Project households. Third, we explore heterogeneity in the peer effect by examining interactions with household characteristics. Finally, we provide a set of robustness checks.

4.1 Primary Results

We begin by estimating our primary specification for water consumption. Table 2 presents the OLS and IV estimates using household and year fixed effects (Columns 1 and 3) and household and year x zip code fixed effects (Columns 2 and 4). These results are robust to substituting house/parcel fixed effects for household fixed effects, where the former is tied to the parcel while latter changes when a house is sold. The primary variable of interest is the log of the mean water consumption of the peers within 500'. Columns 3 and 4 present the two-stage least-squares estimate instrumenting for the peer's water consumption with the fraction of parcels within 500' that observe a housing transaction in that year.

Our preferred specification is in column 4, which indicates that a 10% change in the peer group average water consumption results in a 1.65% or 206.7 gallon per month change in house $i's$ water consumption. To calculate the effect of a change in water use by a single house in the peer group, we can divide the coefficient by the average number of houses in a peer group (25.3): a 10 percentage point change in consumption for a single house in household $i's$ 500' peer group causes a 0.065% change in household $i's$ consumption. For comparison, we estimate that water prices would have to change by 5.22% or $3.22 per month to achieve a similar change in water consumption.[9]

To better understand the spatial nature of this peer effect, we perform a further analysis where we separate the peer group variable into two variables. In one variable, we include only peers within 500' that share the same street as the focal household, while in a second variable we include only those that do not share the same street. Our hypothesis is that individuals are more likely to be influenced by activities that affect water consumption (e.g., changes in landscaping) by close neighbors on the same street than by close neighbors on other streets.

The results are presented in Table 3. Column 1 displays OLS results while column 2 displays IV results. For interpreting the coefficients, it is useful to recognize that the

[9]This calculation was performed using the demand elasticity of -0.33, estimated by Olmstead et al. (2007).

number of houses on the same street is much less than the number of houses not on the same street within the 500' radius of any given house. The bottom two rows of the table show the marginal effect per household; it can be interpreted as the percent change in the log of annual water consumption for the focal household with a 1% change for a single peer household. The results in column 2 suggest that the per-household peer effect is almost twice as large for peers on the same street as peers on other streets, 0.009 and 0.005, respectively. This provides evidence supporting our hypothesis. However, this evidence should only be taken as suggestive, as the per household marginal effects are not statistically different from one another ($p = 0.17$).[10]

The regression results so far provide evidence of peer effects in aggregate water consumption, with visible landscaping as a potential mechanism. To test the hypothesis that landscaping is a primary mechanism for these peer effects, we estimate our model using the satellite landscape data. Table 4 presents the results with household and year fixed effects (Columns 1 and 3) and household and year x zip code fixed effects (Columns 2 and 4). Similar to before, columns 1 and 2 present OLS results, while columns 3 and 4 present the results where we instrument for the peer group's landscape with the fraction of houses in the peer group that has movers. The primary variable of interest, log(percentage green radius 500, is the log of the average percentage of green landscaping for the neighbors within a 500 foot radius of a given household in our sample. Similar to the water consumption regressions (Table 2), the peer effect coefficient decreases slightly in magnitude when time-varying regional unobserveables are controlled for (comparing columns 1 and 2 or 3 and 4). The results from our IV analysis suggests that a 10% increase in the peer group average landscaping results in a 1.39% increase in house $i's$ green landscaping (Column 3). When calculating the per-house peer effect, we find that a 10% increase in green landscaping for a single household in house $i's$ 500 foot peer group causes a 0.06% increase in house $i's$ green landscaping.[11]

[10] We find similar results with the landscape data.

[11] The peer effect coefficient gains statistical significance when year by three digit zip code fixed effects are substituted for year by five digit zip code fixed effects. However, we include year by five digit zip code fixed effects for consistency with the water consumption specifications.

In order to determine the extent to which the water use peer effect is driven by outdoor landscaping decisions, we compare the water savings associated with the peer effect in water consumption and the peer effect in landscape choice. To recap, a 10% change in the peer group average water consumption (average monthly consumption is 12,532 gallons) results in a 1.65% or 206.7 gallon per month change in house $i's$ water consumption with a 10% confidence interval of 119-294 gallons per month. A 10% change in peer landscaping results in a 1.39% change in house $i's$ landscaping. Regressing landscape choice on water consumption (Table A.1) suggests that a 53.2 gallon change in monthly water use (0.42%) results in a 1% change in a household's landscape. Using this conversion, a 10% decrease in peer water use (1,253 gallons per month) applied completely to landscape changes would result in a 3.3% change in house $i's$ landscaping, a savings of 176 gallons per month.[12] This is the same water savings due to peer effects as those calculated from the estimates from the water use regressions. Given that the water consumption peer effect is equivalent, within error bounds, to the landscaping peer effect, it appear that the entirety of the peer effect can be attributed to the conspicuous, outdoor water use.

Finally, we can use the results to estimate the aggregate effect of a stimulus to a focal household (e.g., xeriscaping subsidy) on the peer group. Our results suggest that if a single peer changes landscaping by 1% then the focal household's landscape changes by 0.06%. An intervention to a focal household resulting in a 1% change in outdoor landscaping (roughly 1,879 gallons/month) would result in changes in each peers' landscaping of 0.06%, which is roughly equivalent to an 8.77 gallon change in water consumption per month. On average there are 23.2 peers in a peer group. Therefore, the effect of a stimulus to one household results in 23.2 houses changing their landscape by 0.06% on average, which is equivalent to 203 gallons of water saved per month across the peers. Thus, the spillover effect of the stimulus is equivalent to 11% of the impact on the focal household. Using these numbers, a xeriscaping subsidy that causes 100 distributed households within Phoenix to decrease green landscaping by 1% would save approximately 208,200 gallons of water per month, or roughly 16.5 households monthly water use.

[12]This translates to a savings of $0.81 per month or $9.74 per year per household.

4.1.1 Role of Economic Incentives

If economic savings are important for the operation of the peer effect, we would expect that households which are eligible to receive reduced flood irrigation water (the on-project houses) will be less influenced by peer water consumption and landscape choices. Table 5 provide the results of estimating our preferred specification for off and on project households using the landscaping data. We find that the peer effect is significant and substantial for off-project households, with a coefficient of 0.29, and indistinguishable from zero for on-project households. The difference in the two coefficient estimates is significant at 1%.

This finding suggest that economic incentives to adopt dry landscaping are essential for the operation of peer effects. In neighborhoods where green landscaping is costly, neighbors may discuss xeriscaping (landscaping with slow-growing, drought-tolerant plants) as a money-saving tool. It could be that households are more susceptible to peer effects when there are economic incentives to transition to more arid landscapes, or that peers are less reluctant to share information when there is an obvious monetary benefit. We can not disentangle the two since we don't have a direct measure on the amount of word-of-mouth itself.

Of course, while monetary incentives appear to play a critical role in explaining the difference between off and on project neighborhood peer effects, additional demographic and neighborhood characteristics are also plausible contributing factors to this difference. However, regressing project eligibility on demographic and household characteristics yields very little statistical difference between the two groups.[13]

4.1.2 Heterogeneity in the Effect

In a similar vein, we examine how various demographic and household characteristics influence the peer effect and how demographics correlate with landscape choice. Columns 1, 2, and 3 of Table 6 show the results of our landscaping specification (Column 3 of Table

[13] See Table A.7 in Appendix C.

4) restricted to houses with children, houses with a dog, or houses with a cat respectively. The peer effect is statistically significant for houses with children. On the other hand, we find no evidence of peer effects for households with a dog or a cat, which could be explained if pet-owners have a higher value for green spaces relative to non-pet owners, and are therefore not influenced by peer decisions.[14]

Next we model the impact of neighboring parks and schools on the peer effect. Column 4 of Table 6 provides these results. We find that the peer effect increases with the distance from a park or greenspace. A potential explanation is that households further from parks and greenspaces have less access to publicly available green spaces and are therefore more cognizant of their neighbors' landscape choices. Finally, we model the impact of all of our demographic and household characteristics on peer effects by interacting each of the variables in Table A.3 with the peer variable (percentage green radius 500) and the instrument. With these added controls, we find that the effect of distance to the nearest school mimics the effect of distance to the nearest park. We also see much larger peer effects for households with pools.

To examine how various demographic and household characteristics directly influence landscape choice, we regress the predicted household fixed effects from our main specification (Column 3 of Table 4) on our set of demographic and household variables. The results are presented in Table 7. Column 1 and 2 present the results for off project and on project houses respectively, while column 3 presents the results for the full sample. For off-project households, we find that households that are further from the nearest school and park prefer less green landscaping. It should be noted that these are correlational relationships, it could be that housing costs play a role in this relationship. We find the opposite for on-project households: households prefer *more* green landscaping when they are further from the nearest school and park. This suggests that public green spaces and schools could be substitutes for private green spaces, but that it is only observed when there are low costs for private greenspace.

[14]It should be noted that making these restrictions reduces the sample size by 53%, 82%, and 89% for Columns 1, 2, and 3 respectively.

Dog and cat ownership also leads to more private green space, which is understandable given the animals' needs. It also helps explain the absence of the peer effects for these pet owners, since their pets' needs trump the economic savings from landscape transitions. We also find that higher income households, larger lots, and home value all lead to more green space. Green landscaping is also positively correlated (in the full sample) with being a democrat, having an interest in the outdoors, travel, reading, cooking, or home improvement, and having an interest in exercise, investing, or music is negatively correlated with green landscaping.

To put these result into context, we compare the effect of peers on landscape choices to the effects of owning a dog or living near public green spaces on landscape choices. To do so, we use the regression results from Table 7. We find that the presence of a dog is associated with a 0.2 percentage point increase in green landscaping or an approximate 8% increase in monthly water consumption (1,000 gallons per month). Likewise, being 1,000 feet further from a public green space is associated with a 0.03 percentage point decrease in green landscaping. This comparison suggests that effect of a 1% change in peer landscaping has a larger effect than being 1,000 feet further from public green spaces but has less of an impact than owning a dog.

4.2 Robustness Checks

In this section we provide robustness checks including alternative peer group definitions and a comparison of datasets. Tables 8 and 9 present our preferred specifications using three alternative peer groups defined by 400, 600, and 700 foot radii for water consumption and landscape choice respectively. For both datasets, we find that the peer effect increases with the size of the radius. At first glance these results may seem counterintuitive. However, the peer effect coefficients presented in Tables 8 and 9 represent the effect of changes in the peer group average water consumption and the peer group average landscaping. As the radius of the peer group increases, so do the number of houses contributing to the average. Therefore, as the peer group size increases, in order to achieve

equivalent changes in the peer group average, individual households in the peer group must make larger changes in consumption or landscaping, or an increasing number of households in the peer group must make incremental changes.

Dividing the coefficients by the number of households in each peer group provides the effect of an individual household. For Table 8, a 10% change in water consumption for a single household in $i's$ 400, 600, and 700 foot peer group causes a 0.064%, 0.063%, or 0.063% change in house $i's$ water consumption respectively. For Table 9, a 10% change in green landscaping for a single household in house $i's$ 400, 600, and 700 foot peer group causes a 0.05%, 0.07%, or 0.08% increase in house $i's$ green landscaping respectively. Interestingly, the peer effect of an individual household appears to be relatively constant with distance. Note the standard errors render the coefficients statistically indistinguishable from one another.[15]

One potential concern that might lead to a difference in results for the landscaping and water use analyses is the fact that the houses in the remote sensing sample are a subset of the households in the water consumption sample. Sample selection bias in the landscaping data is therefore a natural concern. Table A.5 presents the results of estimating the models in Table 2, our preferred water consumption regressions, on the remote sensing sample of households. The estimates are statistically indistinguishable across samples indicating the remote sensing sample does not suffer from sample selection bias.

5 Conclusions

In this paper, we estimate causal peer effects in residential water consumption using a unique instrumental variable strategy that leverages peer migration. First, to address simultaneity, we exploit variation in housing sales to develop an instrument for peer landscape choices. Our identifying assumption is that a focal household's landscape choice is

[15]We extended the radius to 1000 feet and found that we lose statistical significance around a radius of 800 feet. Likewise, peer groups defined by radii smaller than 300 feet do not have a statistically significant effect because they contain too few houses.

not directly influenced by housing sales in that household's neighborhood. On the other hand, when individuals move into a new house, they are more likely to make landscape changes at the time of purchase. Second, we use household and year by zip code fixed effects to control for homophily and correlated unobservables.

We then use machine learning techniques on high resolution satellite data, and combine these data with the water billing data to get a robust measure of landscape change. We find that a 10% change in peer water consumption results in a 1.65% change in a focal household's water consumption and a 10% change in peer landscaping results in a 1.39% change in a focal household's landscaping. These results translate to a savings of roughly 200 gallons or roughly $0.81 per household per month, or 2,112 gallons or $9.74 per household per year. We find that the peer effect in landscape transitions can fully explain the peer effect in total water consumption, providing evidence that it is the conspicuous nature of outdoor water conservation that leads to the peer effect.

We further test boundary conditions for the peer effect. We find that the peer effect is stronger for peers on the same street, and that it only operates when there are economic incentives involved. Other factors that contribute to the utility of private green space such as dog and cat ownership can also remove the peer effect completely. The boundary conditions and factors that lead to heterogeneity in the peer effect have important policy implications. For example, targeted conservation programs will have spillover, and potentially multiplying effects. Our research, combined with passed work, suggests the presence of social multipliers in the neighborhood setting (Bollinger and Gillingham 2012). On average, we estimate a targeted landscaping campaign, such as a xeriscaping subsidy, would have a spillover effect equivalent to 11% of impact on the focal household. Furthermore, a xeriscaping subsidy that causes 100 distributed households within Phoenix to decrease green landscaping by 1% would save approximately 208,200 gallons of water per month, or roughly 16.5 households monthly water use. However, such spillovers will only exist in areas with financial incentives to save water and in areas with households that exhibit peer effects.

There are several limitations of this paper worth noting. First, although the images used to construct a measure green landscaping are some of the highest resolution available, they consist of only three color spectrum bands. In order to generate a more robust measure of vegetation, the infrared band is needed. The infrared band allows the computer to see water evaporating from plants, which would allow the computer to better separate green surface such as artificial turf from real grass. Second, more sophisticated software could differentiate between front and back yard vegetation. Having such variation could influence the power of the results as back yard vegetation is not visible to passersby.

By analyzing peer effects for both total water use and landscaping, we are the first to assess the role of visibility for peer effects in conservation behavior. We are also the only study off which we are aware to demonstrate the importance of financial motivation for the operation of peer influence in conservation behavior. Finally, the use of peer migration as an instrument to overcome the challenges associated with identifying peer effects in non-experimental settings will hopefully lead to other similar identification approaches in which exogenous entry into the peer group facilities identification.

References

Allcott, H. (2011), 'Social norms and energy conservation', *Journal of Public Economics* **95**(9-10), 10821095.

Allcott, H. and Rogers, T. (2014), 'The short-run and long-run effects of behavioral interventions: Experimental evidence from energy conservation', *American Economic Review* **104**(10), 3003–3037.
URL: *https://www.aeaweb.org/articles?id=10.1257/aer.104.10.3003*

Angrist, J. D. (2014), 'The perils of peer effects', *Labour Economics* **Volume 30**, 98108.

Arzaghi, M. and Henderson, J. V. (2007), 'Networking off Madison Avenue', *Review of Economic Studies* **75**(4), 1011–1038.

Ayres, I., Raseman, S. and Shih, A. (2012), 'Evidence from two large field experiments that peer comparison feedback can reduce residential energy usage', *Journal of Law, Economics, and Organization* **29**(5), 992–1022.

Bass, F. M. (1969), 'A new product growth model for consumer durables', *Management Science* **15**, 215–227.

Bell, D. R. and Song, S. (2007), 'Neighborhood effects and trial on the internet: Evidence from online grocery retailing', *Quantitative Marketing and Economics* **5**, 361–400.

Bertrand, M., Luttmer, E. F. P. and Mullainathan, S. (2000), 'Network effects and welfare cultures', *Quarterly Journal of Economics* **115**(3), 1019–1055.

Bollinger, B. (2015), 'Green technology adoption: An empirical study of the southern califonnia garment cleaning industry', *Quantitative Marketing and Economics* **13**(4), 319358.

Bollinger, B. and Gillingham, K. (2012), 'Peer effects in the diffusion of solar photovoltaic panels', *Marketing Science* **31**(6), 900–912.

Brock, W. and Durlaf, S. (2001), *Handbook of Econometrics*, Vol. 5, Elsevier, chapter 4, Interaction-based Models, pp. 3297–3380.

Burkhardt, J., Gillingham, K. and Kopelle, P. (2017), 'How do households respond to critical peak pricing? experimental evidence on the role of information and incentives', *Yale University Working Paper* .

Choi, J., Huiand, S. K. and Bell, D. R. (2010), 'Spatiotemporal analysis of imitation behavior across new buyers at an online grocery retailer', *Journal of Marketing Research* **47**(1), 75–89.

Conley, T. and Udry, C. (2010), 'Learning about a new technology: Pineapple in Ghana', *American Economic Review* **100(1)**, 35–69.

Costa, D. L. and Kahn, M. E. (2013), 'Energy conservation nudges and environmentalist ideology: Evidence from a randomized residential electricity field experiment', *Journal of the European Economic Association* **11**, 680702.

Deschnes, O. and Greenstone, M. (2007), 'The economic impacts of climate change: Evidence from agricultural output and random fluctuations in weather', *The American Economic Review* **97**(1), 354–385.

Dolan, P. and Metcalfe, R. (2015), Neighbors, knowledge, and nuggets: Two natural field experiments on the role of incentives on energy conservation.

Ferraro, P. J. and Price, M. K. (2013), 'Using nonpecuniary strategies to influence behavior: Evidence from a large-scale field experiment', *The Review of Economics and Statistics* **95**(1), 64–73.

Fisher, A. C., Hanemann, W. M., Roberts, M. J. and Schlenker, W. (2012), 'The economic impacts of climate change: Evidence from agricultural output and random fluctuations in weather: Comment', *The American Economic Review* **102**(7), 3749–3760.

Foster, A. and Rosenzweig, M. (1995), 'Learning by doing and learning from others: Human capital and technical change in agriculture', *Journal of Political Economy* **103**(6), 1176–1209.

Graham, B. (2008), 'Identifying social interactions through conditional variance restrictions', *Econometrica* **76(3)**, 643–660.

Graziano, M. and Gillingham, K. (2015), 'Spatial patterns of solar photovoltaic system adoption: The influence of neighbors and the built environment', *Journal of Economic Geography* **forthcoming**.

Griliches, Z. (1957), 'Hybrid corn: An exploration in the economics of technological change', *Econometrica* **25**(4), 501–522.

Hartmann, W. R., Manchanda, P., Nair, H., Bothner, M., Dodds, P., Godes, D., Hosanagar, K. and Tucker, C. (2008), 'Modeling social interactions: Identification, empirical methods and policy implications', *Marketing Letters* **19**, 287–304.

Heutel, G. and Muehlegger, E. (2015), 'Consumer learning and hybrid vehicle adoption', *Environmental and Resource Economics* **62**(1), 125–161.
URL: *https://link.springer.com/article/10.1007/s10640-014-9819-3*

Ito, K., Ida, T. and Tanaka, M. (2017), 'Moral suasion and economic incentives: Field experimental evidence from energy demand', *American Economic Journal–Economic Policy* **forthcoming**.

Jessoe, K. and Rapson, D. (2014), 'Knowledge is (less) power: Experimental evidence from residential energy use', *American Economic Review* **104**(4), 1417–1438.
URL: *https://www.aeaweb.org/articles?id=10.1257/aer.104.4.1417*

Manchanda, P., Xie, Y. and Youn, N. (2008), 'The role of targeted communication and contagion in product adoption', *Marketing Science* **27**(6), 961–976.

Manski, C. (1993), 'Identification of endogenous social effects: The reflection problem', *Review of Economic Studies* **60**, 531–542.

McShane, B., Bradlow, E. T. and Berger, J. (2012), 'Visual influence and social groups', *forthcoming at Journal of Marketing Research* .

Moffitt, R. (2001), *Social Dynamics*, MIT Press, chapter 3, Policy Interventions, Low-level Equilibria, and Social Interactions, pp. 45–82.

Nam, S., Manchanda, P. and Chintagunta, P. K. (2010), 'The effects of service quality and word-of-mouth on customer acquisition, retention and usage', *Marketing Science* **29**(4), 690–700.

Narayanan, S. and Nair, H. (2013), 'Estimating causal installed-base effects: A bias-correction approach', *Journal of Marketing Research* **50**(1), 70–94.

Olmstead, S. M., Hanemann, W. M. and Stavins, R. N. (2007), 'Water demand under alternative price structures', *Journal of Environmental Economics and Management* **54**(2), 181 – 198.
URL: *http://www.sciencedirect.com/science/article/pii/S0095069607000605*

Reiss, P. and White, M. (2008), 'What changes energy consumption? prices and public pressures', *RAND Journal of Economics* **39**(3), 636–663.
URL: *https://www.aeaweb.org/articles?id=10.1257/aer.104.10.3003*

Rogers, E. (1995), *Diffusion of Innovations*, The Free Press.

Sacerdote, B. (2001), 'Peer effects with random assignment: Results for Dartmouth roommates', *Quarterly Journal of Economics* **116**, 681–704.

Topa, G. (2001), 'Social interactions, local spillovers and unemployment', *Review of Economic Studies* **68**(2), 261–295.

Towe, C. and Lawley, C. (2013), 'The contagion effect of neighborhood foreclosures', *American Economic Journal: Economic Policy* **5**(2), 313–335.
URL: *http://www.jstor.org/stable/43189336*

Tsvetanov, T. and Gillingham, K. (2017), 'Nudging energy efficiency audits: Evidence from a field experiment', *Yale University Working Paper* .

Tables & Figures

Table 1: Summary Statistics

Variable	Mean	Std. Dev.	Min.	Max.
Panel A: Water Consumption (N=2,533,715; 316,714 households)				
household annual consumption	177.4	111.6	15.0	718.0
annual consumption of households within 500′	177.0	63.9	16.0	717.0
1(housing transaction)	0.045	0.207	0	1
fraction of houses sold within 500′	0.045	0.051	0	1
Panel B: Landscape (N=538,963; 67,370 households)				
fraction of landscape green	0.37	0.10	0.19	0.68
fraction green for households within 500′	0.37	0.06	0.19	0.68
1(housing transaction)	0.04	0.20	0	1
fraction of households sold within 500′	0.04	0.05	0	1

Notes: An observation is a household-year for both datasets. The units of household annual water consumption are in hundreds of gallons per year. Figure 2 illustrates the parcels covered in the landscape subsample and Appendix B provides evidence on the representativeness of this sample. Summary statistics for other radii besides 500′ are also given in Appendix B.

Table 2: Primary Specifications for Water Consumption

	(1) OLS	(2) OLS	(3) IV	(4) IV
log(consumption within 500′)	0.18***	0.12***	0.31***	0.17***
	(0.01)	(0.01)	(0.03)	(0.04)
1(housing transaction)	-0.23***	-0.22***	-0.23***	-0.22***
	(0.003)	(0.003)	(0.003)	(0.003)
Household Fixed Effects	Y	Y	Y	Y
Year Fixed Effects	Y	N	Y	N
Year x Zip Code Fixed Effects	N	Y	N	Y
R-squared	0.77	0.77	0.77	0.77
N	2,533,715	2,533,715	2,533,715	2,533,715

Notes: The dependent variable is the log of annual water consumption. An observation is a household parcel-year. log(consumption within 500′) is the log of the average annual consumption of the neighbors within a 500-foot radius. On average, there are 25.3 houses within a 500′ radius of any household in our study. Column 1 and 2 present OLS peer effect results. Columns 3 and 4 instrument for log(consumption within 500′) using the fraction of parcels with housing transactions within 500′. Standard errors are clustered at the subdivision level. *** denotes significance at 1% level, ** at 5% level, * at 10% level.

Table 3: Street Level Analysis for Water Consumption

	(1) OLS	(2) IV
log(same street peer consumption)	0.054***	0.067***
	(0.004)	(0.023)
log(other street peer consumption)	0.051***	0.091***
	(0.005)	(0.025)
1(housing transaction)	-0.25***	-0.25***
	(0.004)	(0.004)
Household Fixed Effects	Y	Y
Year x Zip Code Fixed Effects	Y	Y
R-squared	0.78	0.78
N	1,959,519	1,959,519
Marginal effect per household		
Same street peer consumption	0.007	0.009
Other street peer consumption	0.003	0.005

Notes: The dependent variable is the log of annual water consumption. An observation is a household parcel-year. log(same street peer consumption) is the log of the average annual consumption of the neighbors within a 500-foot radius who are on the same street as the focal household. log(other street peer consumption) refers to those not on the same street. Column 2 instruments for the same street log consumption using the fraction of parcels with housing transactions within 500' that are on the same street, while the other street log consumption is instrumented with the fraction of parcels with transactions on other streets. On average, there are 7.3 houses on the same street within a 500' radius of any house in our study and 17.5 houses not on the same street within a 500' radius of any house. The final two rows show the calculated marginal effect per household for same and other street peers. Standard errors are clustered at the subdivision level. *** denotes significance at 1% level, ** at 5% level, * at 10% level.

Table 4: Primary Specifications for Landscape Data

	(1) OLS	(2) OLS	(3) IV	(4) IV
log(% green within 500')	0.018***	0.018***	0.315***	0.140*
	(0.003)	(0.003)	(0.105)	(0.073)
1(housing transaction)	-0.007***	-0.007***	-0.007***	-0.007***
	(0.0003)	(0.0003)	(0.0003)	(0.0003)
Household Fixed Effects	Y	Y	Y	Y
Year Fixed Effects	Y	N	Y	N
Year x Zip Code Fixed Effects	N	Y	N	Y
R-squared	0.04	0.08	0.02	0.03
N	538,963	538,963	531,650	531,650

Notes: The dependent variable in each regression is the log of the percentage of green in a given parcel. log(% green within 500'), the variable of interest, is the log of the average percentage of green of the neighbors within a 500' radius. On average, there are 23.2 houses within a 500 foot radius of any household in our study. Column 1 and 2 present OLS results. Columns 3 and 4 instrument the peer variable using the fraction of houses in the peer group that had a housing transaction. Standard errors are clustered at the subdivision level. *** denotes significance at 1% level, ** at 5% level, * at 10% level.

Table 5: Landscape Preferred Specifications for Off/On Project Households

	(1) Off Project	(2) On Project
log percentage green radius 500	0.2887***	-0.0581
	(0.1044)	(0.0839)
move this year	-0.0070***	-0.0064***
	(0.0004)	(0.0003)
Household Fixed Effects	Y	Y
Year X Zip Fixed Effects	Y	Y
R-squared	0.026	0.034
N	328620	203030

Notes: Preferred IV specification for on/off project households. Standard errors are clustered at the subdivision level.
***Significance at the 1 percent level.
**Significance at the 5 percent level.
*Significance at the 10 percent level.

Table 6: Landscape Demographic Restrictions

	(1) has kids	(2) has dog	(3) has cat	(4) interactions	(5) interactions
log percentage green radius 500	0.2474**	0.0369	-0.1052	0.0979	-0.3943
	(0.1097)	(0.1750)	(0.2098)	(0.0910)	(0.3452)
move this year	-0.0076***	-0.0076***	-0.0072***	-0.0075***	-0.0073***
	(0.0003)	(0.0006)	(0.0008)	(0.0003)	(0.0003)
peer X park				0.0429***	0.0595***
				(0.0163)	(0.0200)
peer X school				0.0329	0.0778*
				(0.0329)	(0.0423)
peer X kids					0.0332
					(0.0445)
peer X cat owner					-0.0543
					(0.0900)
peer X dog owner					0.0628
					(0.0907)
peer X age					0.0038
					(0.0040)
peer X home value					-0.0016
					(0.0153)
peer X pool					0.2222***
					(0.0802)
peer X sports					-0.0160
					(0.0971)
peer X outdoors					0.0217
					(0.0852)
peer X travel					-0.0435
					(0.0957)
peer X reading					0.0787
					(0.1259)
peer X cooking					-0.1597
					(0.1331)
peer X exercise					-0.1085
					(0.1116)
peer X music					-0.0617
					(0.1104)
peer X electronics					0.1966*
					(0.1165)
peer X home improvement					-0.0990
					(0.1197)
peer X investing					0.0772
					(0.0848)
peer X antiques					-0.0719
					(0.0923)
Household Fixed Effects	Y	Y	Y	Y	Y
Year Fixed Effects	Y	Y	Y	Y	Y
R-squared	0.009	0.042	0.033	0.010	0.015
N	251001	99619	57679	538631	339030

Notes: Columns 1-3 restrict the sample to houses with children, a dog, or a cat respectively. Columns 4 and 5 include interactions between the demographic variables and the peer green variable. Standard errors are clustered at the subdivision level.
***Significance at the 1 percent level.
**Significance at the 5 percent level.
*Significance at the 10 percent level.

Table 7: Landscape Regression Fixed Effects on Demographics

	(1) off project	(2) on project	(3) full sample
distance to school (1,000 ft)	-0.0005***	0.0032***	-0.0006***
	(0.0002)	(0.0003)	(0.0002)
distance to park (1,000 ft)	-0.0011***	0.0014***	-0.0003***
	(0.0001)	(0.0002)	(0.0001)
lot size (1,000 sq ft)	1.4679***	7.9256***	2.1484***
	(0.0340)	(0.0879)	(0.0317)
income bracket	0.0012***	0.0018***	0.0016***
	(0.0001)	(0.0001)	(0.0001)
age	-0.0001***	0.0001***	0.0000
	(0.0000)	(0.0000)	(0.0000)
home value	0.0019***	0.0034***	0.0034***
	(0.0001)	(0.0001)	(0.0001)
number of kids	-0.0003	-0.0021***	-0.0009***
	(0.0002)	(0.0003)	(0.0002)
house size (1,000 sq ft)	2.7294***	-21.0662***	-0.3607
	(0.4288)	(0.6629)	(0.3580)
construct	-0.001***	-0.001***	-0.001***
	(0.000)	(0.000)	(0.000)
has pool	-0.0141***	-0.0234***	-0.0175***
	(0.0004)	(0.0006)	(0.0003)
has cat	0.0065***	0.0053***	0.0061***
	(0.0005)	(0.0008)	(0.0005)
has dog	0.0032***	0.0002	0.0020***
	(0.0005)	(0.0008)	(0.0004)
renter	0.0053***	-0.0041**	0.0022*
	(0.0016)	(0.0019)	(0.0012)
democrat	0.0017***	0.0014**	0.0020***
	(0.0004)	(0.0006)	(0.0003)
group sports	-0.0013**	0.0049***	0.0008*
	(0.0005)	(0.0007)	(0.0004)
group outdoors	0.0020***	0.0002	0.0009**
	(0.0005)	(0.0007)	(0.0004)
group travel	0.0032***	0.0009	0.0025***
	(0.0005)	(0.0007)	(0.0004)
group reading	-0.0002	0.0077***	0.0033***
	(0.0007)	(0.0009)	(0.0006)
group cooking	0.0079***	0.0072***	0.0077***
	(0.0006)	(0.0008)	(0.0005)
group exercise	-0.0034***	-0.0019**	-0.0032***
	(0.0006)	(0.0008)	(0.0005)
group music	-0.0034***	-0.0050***	-0.0042***
	(0.0005)	(0.0008)	(0.0004)
group electronics	0.0034***	0.0019**	0.0029***
	(0.0007)	(0.0008)	(0.0005)
group home improvement	0.0031***	0.0069***	0.0048***
	(0.0006)	(0.0009)	(0.0005)
group investing	-0.0013***	-0.0055***	-0.0034***
	(0.0005)	(0.0007)	(0.0004)
group antiques	0.0006	0.0009	0.0009**
	(0.0004)	(0.0007)	(0.0004)
group SOHO	0.0029***	-0.0008	0.0013**
	(0.0006)	(0.0009)	(0.0005)
R-squared	0.043	0.185	0.097
N	215557	123684	339241

Notes: Results of regressing post estimation fixed effects on demographics.
***Significance at the 1 percent level.
**Significance at the 5 percent level.
*Significance at the 10 percent level.

Table 8: Water Consumption: Comparing Different Radii

	(1)	(2)	(3)
log consumption radius 400	0.1411***		
	(0.0341)		
log consumption radius 600		0.1652***	
		(0.0371)	
log consumption radius 700			0.1689***
			(0.0371)
move this year	-0.2240***	-0.2239***	-0.2239***
	(0.0031)	(0.0031)	(0.0031)
Household Fixed Effects	Y	Y	Y
Year X Zip Fixed Effects	Y	Y	Y
R-squared	0.771	0.772	0.772
N	2532336	2534459	2534922

Notes: Results of preferred IV specification for alternative radii. On average, there are respectively 22.9, 26.25, and 26.7 houses within a 400, 600, and 700 foot radius of any household in our study. Standard errors are clustered at the subdivision level.
***Significance at the 1 percent level.
**Significance at the 5 percent level.
*Significance at the 10 percent level.

Table 9: Landscape: Comparing Different Radii

	(1)	(2)	(3)
log percentage green radius 400	0.0984		
	(0.0651)		
log percentage green radius 600		0.1889*	
		(0.0739)	
log percentage green radius 700			0.2159**
			(0.0797)
move this year	-0.0067***	-0.0067***	-0.0067***
	(0.0003)	(0.0003)	(0.0003)
Household Fixed Effects	Y	Y	Y
Year X Zip Fixed Effects	Y	Y	Y
R-squared	0.032	0.023	0.019
N	531461	531704	531759

Notes: Preferred IV specification for alternative radii. On average, there are respectively 19.7, 24.9, and 25.6 houses within a 400, 600, and 700 foot radius of any household in our study. Standard errors are clustered at the subdivision level.
***Significance at the 1 percent level.
**Significance at the 5 percent level.
*Significance at the 10 percent level.

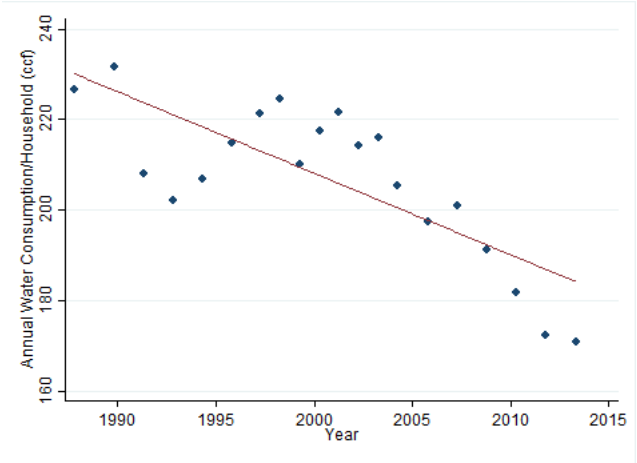


Figure 1: Binscatter plot of the annual water consumption per household in Phoenix along with a linear best-fit trendline. Data source: City of Phoenix Water Services Department.

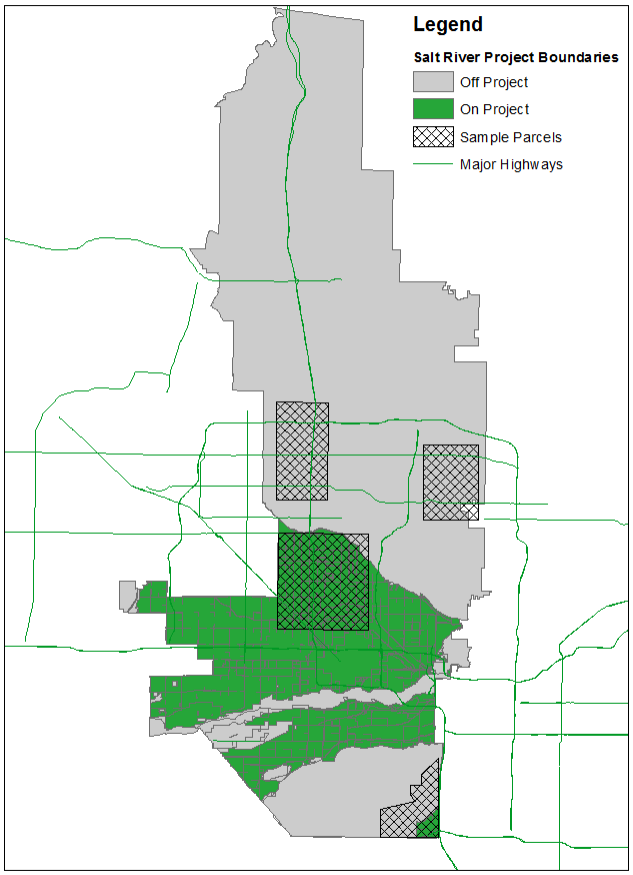


Figure 2: Phoenix water district boundary, along with identification of areas under the Salt River Project and landscape sample parcels.



Figure 3: Illustrative satellite images demonstrating the classification of green space. Panel A on the left shows what our satellite images look like, while Panel B on the left shows how the machine learning algorithm codes the pixels of green space.

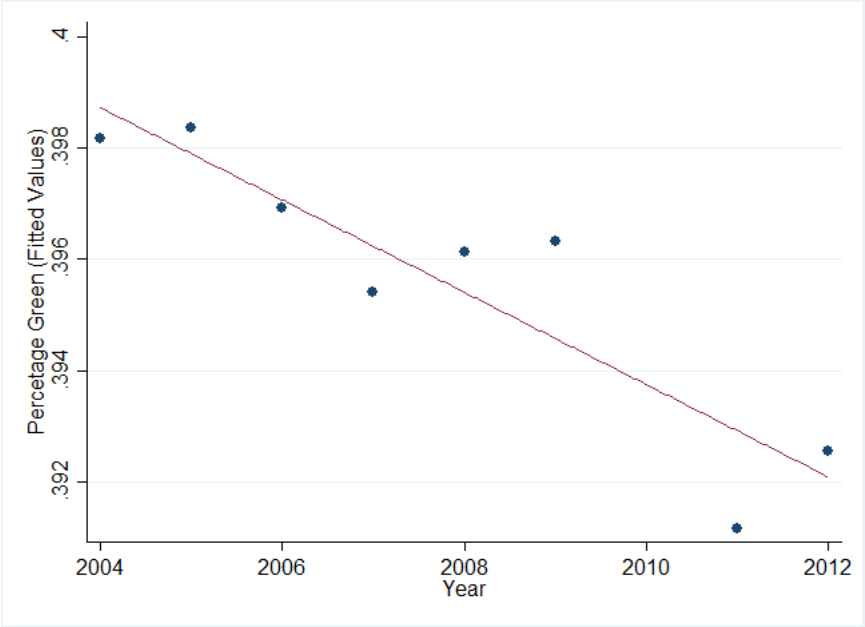


Figure 4: The mean landscape greenness variable by year, based on on the satellite imagery.

Appendix A: Remote Sensing Image Development Methodology

One issue that we must contend with is that remote sensing data contains inherent measurement error from the time of day (and slight difference in the time of year) that the photo is taken, which could be an issue if we cannot control for correlated shocks across individuals.[16] To show this formally, let u_{it} be the measurement error. Thus, we observe:

$$\tilde{y_{it}} = y_{it} + u_{it}, \tag{2}$$

and so:

$$\mathbb{E}\left[\tilde{y_{it}}\right] = \frac{(\alpha + \beta x_{gt} + \gamma w_{gt} + \delta \boldsymbol{m}_{gt})}{1-\theta} + u_{gt}.$$

By substitution we have that:

$$\tilde{y_{it}} = \alpha + \beta x_{it} + \gamma w_{gt} + \theta \mathbb{E}\left[\tilde{y}_{gt}\right] + \delta \boldsymbol{m}_{it} + u_{it} - \theta u_{gt} + \epsilon_{it}. \tag{3}$$

Clearly, there is correlation between the measured landscape of peers and the compound error term, $\theta u_{gt} + \epsilon_{it}$. If there is no correlation between the move variable of peers and the extent of the measurement error, then our instrumental variables strategy will remove the measurement error. For example, if $E[u_{it}\boldsymbol{m}_{it}] = 0$, then we have a valid IV for $\mathbb{E}\left[\tilde{y}_{gt}\right]$ in $\boldsymbol{m}_g$. However, if the measurement error is correlated with characteristics that also correlate to whether peers are more likely to move to certain neighborhoods (such as lot size), so that $E[u_{it}\boldsymbol{m}_{it}|w_g, t] \neq 0$, then the instrumentation strategy by itself will not remove potential bias caused by measurement error. This is analogous to the issue noted by Fisher et al. (2012) regarding Deschnes and Greenstone (2007), in which significant within-state weather variation in their Table 2 was largely an artifact of errors in

[16] Although we use an instrumental variables identification strategy, when we estimate our primary specification using the uncorrected percentage green and peer green variables, the peer effect coefficient is spurious and not well identified. This suggests the measurement error is not fully addressed by simply instrumenting the uncorrected images.

the weather data, which exhibited discontinuous shifts across neighboring counties.[17] To address the issue, Deschnes and Greenstone (2007) recreate the weather variables.

Similarly, to address the remote sensing measurement error issue, we use water consumption data. We know (from basic botany) that the greenness of a lot is a direct function of how much water is used for irrigation. Thus in actuality:

$$\begin{aligned} S_i * y_{it} = \sum_{m \in t} q_{itm}^{out} \psi + \xi_{it}^{out} &= \sum_{m \in t} q_{itm} \psi_m - \sum_{m \in t} \psi_m q_{itm}^{in} + \xi_{it}^{out} \\ &= \sum_{m \in t} q_{itm} \psi_m + \zeta_i + \xi_{it}, \end{aligned} \tag{4}$$

where S_i is the lot size of house i, q_{itm}^{out} is monthly outdoor water usage, q_{itm}^{in} is monthly indoor water usage, q_{itm} is total monthly water usage, and ψ is a household fixed effect. A benign assumption is that the remote sensing image measurement error is not correlated with water usage.

By substitution we can then also write:

$$S_i * \tilde{y}_{it} = \sum_{m \in t} q_{itm} \psi_m + \zeta_i + \xi_{it} + S_i u_{it}, \tag{5}$$

We can estimate $\hat{\psi}_m$ in a first stage using equation (5). We further define the fitted $\hat{y}_{it}$ from the first stage regression as:

$$\hat{y_{it}} \equiv \frac{1}{S_i} \left(\sum_{m \in t} q_{itm} \hat{\psi}_m + \zeta_i \right). \tag{6}$$

Note we can write:

$$y_{it} = \hat{y}_{it} + \frac{1}{S_i} \xi_{it} + e_{it}, \tag{7}$$

where e_{it} is the first stage estimation error.

[17] Deschnes and Greenstone (2007) also remark that their fixed effects regressions will exacerbate issues from misspecification from measurement error.

By substitution of (7) into (1), we have:

$$\hat{y_{it}} = \alpha + \beta x_{it} + \gamma w_{gt} + \theta \mathbb{E}\left[\hat{y_g}\right] + \theta \mathbb{E}\left[\frac{1}{S_i}\xi_{gt}\right] + \delta \boldsymbol{m}_i - \frac{1}{S_i}\xi_{it} - e_{it} + \epsilon_{it}, \tag{8}$$

where $\mathbb{E}\left[\frac{1}{S_i}\xi_{gt}\right]$ is the expected unobserved water usage per area by i's peers. If it is the case that $E[\xi_{it}\boldsymbol{m}_{it}|w_g, t] = 0$, then $E[\xi_g \boldsymbol{m}_g|w_g, t] = 0$ and so $\boldsymbol{m}_{gt}$ is a valid IV for $[\hat{y_{gt}}]$. In practice, we estimate equation 8 separately for on and off Salt River Project households.

There is absolutely no reason why shocks to water consumption would be correlated with measurement error in the remote sensing image data caused by shadows, haze, or weather that depend on the time of day that the remote sensing photo was taken. Thus we are confident that the triangulation of the remote sensing image data and water consumption data provide an accurate measure of actual landscape values.

Table A.1: Data Cleaning First Stage Regressions

	(1) Off Project	(2) On Project	(3) Full Sample
consjan	-1.5068***	-0.5621***	-1.0941***
	(0.1663)	(0.1828)	(0.1232)
consfeb	-3.0540***	-1.5572***	-2.2900***
	(0.2235)	(0.2145)	(0.1554)
consmar	3.0984***	-0.3952	1.7730***
	(0.3963)	(0.4869)	(0.3071)
consapr	0.9188***	0.2869**	0.5829***
	(0.1478)	(0.1348)	(0.1001)
consmay	1.3168***	1.4489***	1.4596***
	(0.2842)	(0.3317)	(0.2153)
consjun	2.9037***	1.9201***	2.5240***
	(0.2813)	(0.2748)	(0.1956)
consjul	-0.0389	1.7344***	0.3114**
	(0.1661)	(0.3338)	(0.1477)
consaug	0.6124***	0.1701	0.5552***
	(0.1536)	(0.2400)	(0.1286)
conssep	-0.9653***	-0.6098***	-0.7721***
	(0.1776)	(0.2242)	(0.1388)
consoct	2.3744***	1.3273***	1.7625***
	(0.2189)	(0.1719)	(0.1360)
consnov	-0.2155	0.8837***	0.5128***
	(0.1592)	(0.1540)	(0.1076)
consdec	3.2669***	1.5089***	2.1375***
	(0.2012)	(0.1565)	(0.1221)
Year Fixed Effects	Y	Y	Y
Household Fixed Effects	Y	Y	Y
R-squared	0.840	0.843	0.841
N	340643	214239	554882

Notes: Dependent variable is quantity of green landscaping in a given parcel in square feet, i.e., lot size X percentage green. Independent variables are monthly water consumption.
***Significance at the 0.1 percent level.
**Significance at the 1 percent level.
*Significance at the 5 percent level.

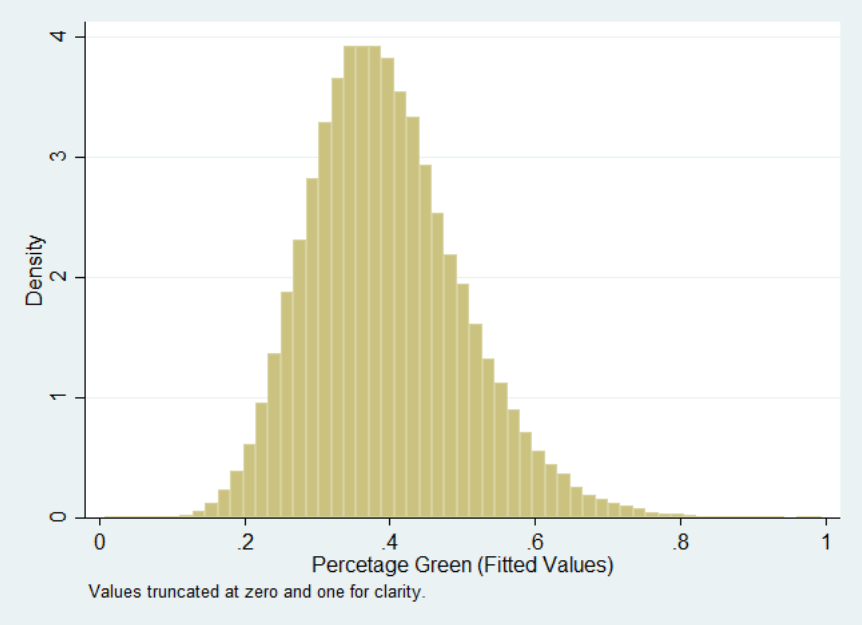


Figure A.1: Histogram of the percentage of the land area of a parcel that is green space, as coded by the machine learning algorithm and adjusted to address remote sensing measurement error.

Appendix B: Further Summary Statistics

This appendix contains a further descriptive figure and summary statistics for some of the variables used in our analysis of heterogeneity of the peer effect. The figure below shows the substantial variation in our landscape variable.

Table A.2: Summary Statistics: Household Characteristics

Variable	Mean	Std. Dev.	Min.	Max.	N
# bathroom fixtures	7.092	2.601	1	40	308215
home sq ft	1791.241	713.236	240	21012	308222
construction date	1978.567	19.172	1900	2012	308222
pool sq ft	439.626	102.204	51	4275	101622
1(SRP project)	0.434	0.496	0	1	308601
lot size (ft^2)	8807.869	6035.955	531.3	222789.719	297744

Appendix C: Validation of Identification Assumption

Table A.3: Summary Statistics: Demographic Variables

Variable	Mean	Std. Dev.	Min.	Max.	N
lot size (1,000 sq ft)	0.01	0.006	0.002	0.211	71802
income bracket	7.861	3.632	1	13	59438
age	56.899	14.941	18	99	45291
home value	10.169	4.155	1	19	58856
number of kids	0.519	0.838	0	6	59438
house size (1,000 sq ft)	0.002	0.001	0	0.01	71625
construction date	1971.271	15.424	1900	2012	71625
has pool	0.42	0.494	0	1	71802
has cat	0.129	0.335	0	1	59439
has dog	0.222	0.415	0	1	59439
renter	0.081	0.273	0	1	59438
democrat	0.231	0.421	0	1	59439
group sports	0.457	0.498	0	1	59439
group outdoors	0.39	0.488	0	1	59439
group travel	0.5	0.5	0	1	59439
group reading	0.621	0.485	0	1	59439
group cooking	0.5	0.5	0	1	59439
group exercise	0.497	0.5	0	1	59439
group music	0.403	0.49	0	1	59439
group electronics	0.635	0.481	0	1	59439
group home improvement	0.566	0.496	0	1	59439
group investing	0.451	0.498	0	1	59439
group antiques	0.344	0.475	0	1	59439
group SOHO	0.086	0.28	0	1	59439
distance to school (1,000 ft)	1.488	1.033	0	8.083	71802
distance to park (1,000 ft)	2.725	1.918	0	10.958	71802

Notes: Distance to school or park may be zero due to how Arcmap defines boundaries of polygons. Demographic variables only apply to remote sensing image dataset.

Table A.4: Test of Identifying Assumption

	(1) Water Consumption	(2) Landscape
Δ_1(peer consumption/landscape)	-0.0039	-0.0807
	(0.0024)	(0.0608)
Δ_2(peer consumption/landscape)	0.0026	-0.0918
	(0.0021)	(0.0696)
R-squared	0.118	0.142
N	1924912	390620

Notes: Dependent variable is an indicator for whether or not a house was sold in a particular year. Column 1 is for the water consumption dataset while column 2 is for the landscape dataset. Water consumption changes are in millions of gallons. Standard errors are clustered at the subdivision level.

***Significance at the 0.1 percent level.

**Significance at the 1 percent level.

*Significance at the 5 percent level.

Appendix D: First Stage Results and Robustness Checks

Table A.5: Water Consumption On Remote Sensing Image Dataset

	(1) OLS	(2) OLS	(3) IV	(4) IV
log consumption radius 500	0.1360***	0.0814***	0.3571***	0.2427***
	(0.0116)	(0.0120)	(0.0740)	(0.0929)
move this year	-0.1786***	-0.1784***	-0.1784***	-0.1783***
	(0.0055)	(0.0055)	(0.0055)	(0.0055)
Household Fixed Effects	Y	Y	Y	Y
Year Fixed Effects	Y		Y	
Year X Zip Fixed Effects		Y		Y
R-squared	0.797	0.798	0.797	0.798
N	533426	533426	533426	533426

Notes: Specifications are the same as Table 2 but estimated on the remote sensing image sample. Standard errors are clustered at the subdivision level.
***Significance at the 1 percent level.
**Significance at the 5 percent level.
*Significance at the 10 percent level.

Table A.6: First Stage Preferred IV for Water and Landscape Data

	(1)	(2)
	log peer consumption	log peer landscape
mean sold houses radius 500	-0.2029***	-0.0096***
	(0.0072)	(0.0012)
move this year	-0.0031***	-0.0000
	(0.0006)	(0.0001)
Household Fixed Effects	Y	Y
Year by Zip Fixed Effects	Y	Y
R-squared	0.947	0.421
N	2533715	538963

Notes: Dependent variable in column 1 is the log of average water consumption for $i's$ peers within a 500 foot radius. Dependent variable in column 2 is log of average green landscaping for $i's$ peers within a 500 foot radius. Mean sold houses radius 500 is the average number of sold houses with $i's$ peer group. Standard errors are clustered at the subdivision level.
***Significance at the 0.1 percent level.
**Significance at the 1 percent level.
*Significance at the 5 percent level.

Table A.7: Project Code on Demographics and Housing Characteristics

	(1) Project Code
lot size (1000 sq ft)	0.0006**
	(0.0003)
number of kids	-0.0002
	(0.0010)
home value	0.0009**
	(0.0005)
income bracket	-0.0003
	(0.0003)
has cat	0.0018
	(0.0027)
has dog	-0.0007
	(0.0024)
group sports	0.0026
	(0.0025)
group outdoors	0.0010
	(0.0024)
group travel	-0.0043
	(0.0027)
group reading	-0.0007
	(0.0034)
group cooking	-0.0013
	(0.0029)
group exercise	-0.0062**
	(0.0027)
group music	0.0033
	(0.0025)
group electronics	-0.0012
	(0.0032)
group home improvement	0.0071**
	(0.0032)
group investing	0.0007
	(0.0023)
group antiques	-0.0047**
	(0.0022)
group SOHO	0.0024
	(0.0029)
renter	0.0030
	(0.0038)
democrat	-0.0006
	(0.0020)
distance to park (100 ft)	-0.0003
	(0.0002)
distance to school (100 ft)	-0.0001
	(0.0002)
construction date	-0.0001
	(0.0003)
garage size (1000 sq ft)	0.0089
	(0.0058)
house size (1000 sq ft)	-0.0022
	(0.0022)
pool size (1000 sq ft)	-0.0166*
	(0.0093)
Subdivision FE	Y
R-squared	0.931
N	22764

Notes: Dependent variable is project code, i.e., project eligibility.
***Significance at the 0.1 percent level.
**Significance at the 1 percent level.
*Significance at the 5 percent level.