Social Comparisons and Groundwater Use: Evidence from Colorado and Kansas *

R. Aaron Hrozencik† Jordan F. Suter‡ Paul J. Ferraro§
Nathan Hendricks¶

November 2, 2022

Abstract

In the United States, agriculture is responsible for the majority of consumptive water use. In an effort to reduce water use in water scarce regions, policymakers have implemented a number of costly interventions. These interventions range from land retirement to subsidies that encourage the adoption of efficient irrigation technologies. In non-agricultural contexts, costly policy interventions have been complemented by low-cost interventions inspired by behavioral economics. Whether these behavioral interventions are effective in the context of commercial farming is not well understood. In a pre-registered, randomized field intervention, we estimate the impact of social (peer) comparisons on agricultural groundwater user in Colorado and Kansas. Over three thousand irrigators were randomized to receive either an annual peer comparison or no comparison. The peer comparison contrasted each irrigator's groundwater use to the distribution of use by neighboring irrigators. The comparison intervention reduced average annual groundwater use by 4.05% [95% CI (-5.87%, -2.21%)], resulting in an aggregate reduction of more than 27,000 acre-feet per year at a cost less than $0.10 per acre-foot conserved. The estimated treatment effect was

*The findings and conclusions in this manuscript are those of the authors and should not be construed to represent any official USDA or U.S. government determination or policy. Ferraro acknowledges support from the United States Department of Agriculture – National Institute of Food and Agriculture (2019-67023-29854; 2017-67024-26278).

†USDA-Economic Research Service, Resource and Rural Economics Division

‡Colorado State University, Department of Agricultural and Resource Economics

§Johns Hopkins University, Carey Business School and Department of Environmental Health and Engineering, a joint Department of the Bloomberg School of Public Health and the Whiting School of Engineering.

¶Kansas State University, Department of Agricultural Economics

larger among irrigators with lower pre-intervention water use. The results imply that social comparisons can be a cost-effective tool, alongside other policy interventions, aimed at reducing agricultural water use.

1 Introduction

To design and implement policy interventions to manage natural resources and environmental goods, scholars and practitioners are increasingly leveraging insights from the behavioral sciences (Carlsson and Johansson-Stenman, 2012; Croson and Treich, 2014; Yoeli et al., 2017; Palm-Forster et al., 2019; Carlsson et al., 2021). In a parallel literature, scholars and practitioners are drawing on lab and field experiments to better understand how economic agents interact with environmental goods and natural resources (Allcott and Mullainathan, 2010; Hahn and Metcalfe, 2016; Andor and Fels, 2018; Slough et al., 2021).

We build on these two literatures by implementing field experiments that test the impact of social comparison messaging on groundwater use in Colorado and Kansas. Inspired by research in social psychology (Festinger, 1954), social comparisons combine injunctive and descriptive normative messages to induce individuals to change their behavior. Injunctive norms refer to perceptions of what ought to be, while descriptive norms refer to perceptions of what is (Cialdini et al., 1990).

The literature on the effectiveness of social comparisons focuses on consumer behaviors across a range of environmental and non-environmental contexts. These contexts include residential water and electricity consumption (Allcott, 2011; Ayres et al., 2012; Bernedo et al., 2014; Brent et al., 2015; Bhanot, 2017; Ferraro et al., 2017; Torres and Carlsson, 2018; Jessoe et al., 2021a), retirement savings (Beshears et al., 2015), towel reuse in hotels (Goldstein et al., 2008), voting (Gerber and Rogers, 2009), and charitable giving (Frey and Meier, 2004; Shang and Croson, 2009). Evidence of the efficacy of social comparison messaging in curbing household water and electricity demand has prompted some regulators to consider social comparison interventions as demand management tools (e.g., Oracle's OPower product).

Much less attention, however, has been paid to understanding how social comparisons may affect producer behaviors (Wallander et al., 2017; Chabé-Ferret et al., 2019; Earnhart and Ferraro, 2020).[1] While consumer behaviors have important effects on the quality and quantity of environmental goods and natural resources, producer behaviors are also important. In particular, agricultural producers play an important role in maintaining water quality, supporting biodiversity, and determining groundwater stocks (Riseng et al., 2011; Famiglietti, 2014; Van Meter et al., 2018; Tsioumani, 2019).

A growing reliance on groundwater to irrigate crops has led to rates of groundwater pumping in excess of natural recharge, causing groundwater depletion in many of the world's most important and productive aquifers (Konikow and Kendy, 2005; Konikow, 2011; Bierkens and Wada, 2019). To curb groundwater overdraft, policy-makers are considering a range of regulations and policies (e.g., pumping taxes, quantity restrictions, irrigated land and well retirement). However, many of these policies impose significant costs on the irrigated agricultural sector or require significant government expenditures, which raises questions regarding the feasibility of widespread implementation (Feinerman and Knapp, 1983; Nieswiadomy, 1985; Guilfoos et al., 2016; Hrozencik et al., 2017; Tsvetanov and Earnhart, 2020; Manning et al., 2020). Taxes and quantity restrictions imposed by local districts can be effective at decreasing water use without a large decrease in irrigated area (Smith et al., 2017; Drysdale and Hendricks, 2018; Deines et al., 2019), but it is unclear that these measures will be widely adopted without significant threats by state regulators (Perez-Quesada and Hendricks, 2021). The potentially negative impacts of continued aquifer depletion paired with the costs associated with traditional policy options motivates a need to evaluate the impact of behavioral interventions in the context of agricultural groundwater use (Edwards and Guilfoos, 2020).

This paper addresses this need by reporting results from a randomized controlled trial (RCT) that tests the effectiveness of social comparison messaging among irrigated agricultural producers in the High Plains (Ogallala) Aquifer (HPA) region of the U.S.

[1] Reeves (2012) and Sacarny et al. (2016) analyze the impact of social comparisons among medical doctors. However, whether the doctors in these experiments are employees (agents) or firms (principals) is unclear.

The HPA is the most intensively pumped aquifer in the U.S. and irrigated agriculture is responsible for 95% of all withdrawals (Lovelace et al., 2020). Withdrawals from the HPA generate significant value for the U.S. agricultural economy. Suárez et al. (2018) estimate that the HPA annually contributes $3 billion dollars to domestic agricultural production. The extensive use of the HPA to support irrigated agriculture has led to large declines in groundwater stocks across the region and the most significant depletion rates among all domestic aquifers used primarily for agricultural purposes (Konikow, 2015). Haacker et al. (2015) predict the end of economically feasible groundwater-fed irrigated agriculture by 2050 for some regions of the HPA.

In a pre-registered, randomized controlled field experiment, over 1,600 irrigators in Colorado and Kansas were randomly selected to receive social comparisons annually for two or three years (link). The comparisons, which were delivered by mail prior to the growing season, compared the recipient's irrigation water use in the previous year to the overall distribution of irrigation water use in the recipient's groundwater management district (Colorado) or county (Kansas). Colorado irrigators received mailers prior to the 2019, 2020, and 2021 growing seasons while Kansas irrigators received mailers prior to the 2020 and 2021 growing seasons.

Using administrative data from well meters, we estimate that recipients of the social comparisons, on average, used 4.05% [95% CI (-5.87%, -2.21%)] less water than non-recipients in the control group. In our pre-registered analysis of heterogeneous treatment effects, we also find larger estimated treatment effects among irrigators with lower pre-intervention water use.

The estimated reduction from the social comparison is modest, but economically significant, implying a reduction of 6.5 acre-feet per year for the median irrigator and an aggregate annual reduction in groundwater use of more than 27,000 acre-feet, at a cost less than $0.10 per acre-foot conserved. That cost is orders of magnitude less than other groundwater conservation policies implemented in the study area. For example, Tsvetanov and Earnhart (2020) report that the annual program costs of a well retirement program in Kansas were $17.26 per acre-foot conserved.

The cost-effectiveness of the social comparisons in our study highlights how interventions inspired by behavioral economics may have a role alongside traditional agricultural and environmental policies. More broadly, our results suggest that interventions that have been reported to influence consumer behavior may also be effective in influencing producer behavior. As documented by Ferraro et al. (2022), behavioral experiments in which the units of randomization are producers (i.e., firms) operating in competitive environments are extremely rare.

Our study also advances the behavioral and experimental literature by using large sample sizes from multiple regions of the country. In economics (Ioannidis et al., 2017), including environmental economics (Ferraro and Shukla, 2020) and agricultural economics (Ferraro and Shukla, 2022), empirical designs with insufficient statistical power are widespread and can lead to both undetected and exaggerated estimated effect sizes. Achieving sufficient statistical power to detect the small effect sizes reported in the social comparison literature is particularly challenging (Allcott, 2011; Ayres et al., 2012; Bernedo et al., 2014; Brent et al., 2015; Bhanot, 2017; Ferraro et al., 2017; Torres and Carlsson, 2018; Jessoe et al., 2021a). For example, our study is most closely related to Chabé-Ferret et al. (2019), who assess the impacts of a social comparison intervention implemented among French irrigators receiving surface water from a regional supplier. They fail to detect a difference between treatment and control groups, but their sample size of 200 irrigators is insufficient to precisely estimate small, but economically relevant, differences. In contrast, our study with more than 3,000 irrigators was designed to detect effect sizes as small as 3% with 80% power (5% Type 1 error rate). Our large sample size from multiple regions also helps address a common criticism regarding lab and field experiments: weak external validity. This criticism posits that insights generated by experiments often do not generalize to other settings with different attributes

(Roe and Just, 2009; Deaton, 2010; Barrett and Carter, 2020; Ferraro and Agrawal, 2021). By implementing our experiment across multiple counties in two states with differing institutional and environmental attributes, we mitigate this criticism.

The paper is organized as follows. Section 2 provides additional context for both the Colorado and Kansas study areas. Sections 3 and 4 discuss the design of the field experiment and present the econometric model used to assess the impact of the intervention. Finally, sections 5 and 6 present the results of the experiment and discuss their significance to the literature and to broader efforts to promote natural resource conservation and stewardship.

2 Background

The HPA is a vast groundwater resource underlying eight states. Pumping from the HPA accounts for nearly 15% of all groundwater use in the U.S. (Lovelace et al., 2020). The HPA is divided by the 100^{th} meridian that historically separates the humid east from the arid west. That division generates varying climatic conditions along an east-west gradient. The easternmost sections of the HPA in Kansas and Nebraska receive, on average, in excess of 16 inches of precipitation throughout the growing season (PRISM, 2021). Meanwhile, in the western regions of the aquifer, such as Colorado and Wyoming, there is generally less than 10 inches of precipitation during the growing season. The current and predicted future availability of groundwater resources varies significantly across the HPA. Some regions are predicted to reach depletion levels not able to sustain groundwater-fed irrigated agriculture by 2050, while other regions have sufficient groundwater stocks to support irrigation for another thousand years or more (Haacker et al., 2015; Steward and Allen, 2016).

The depletion of the HPA has prompted growing interest in managing the region's shared groundwater resources. Interventions to address the depletion of the HPA have varied in approach and efficacy. Early efforts focused on subsidizing the adoption of efficient irrigation technology. However, the effectiveness of these efforts in curbing water use has been questioned by Pfeiffer and Lin (2014) who find that increases in irrigation efficiency do not necessarily generate water conservation as producers respond to increased application efficiency by adjusting along other margins (e.g., irrigating more land). More recently, groundwater conservation efforts have focused on implementing pumping quantity restrictions and well retirement programs. For example, some regions of Kansas and Nebraska have enacted pumping quantity restrictions (Montginoul et al., 2016; Drysdale and Hendricks, 2018). While these efforts have proven effective in reducing groundwater overdraft, pumping quantity restrictions remain relatively scarce throughout the HPA. The scarcity of groundwater pricing and use regulations may relate to the relatively high economic and political costs imposed by these policies on the irrigated agricultural sector and policy-makers. Finally, federal and state governments have jointly funded well retirement programs in the HPA region to pay agricultural producers to take wells and irrigated lands out of production (Monger et al., 2018; Rosenberg, 2020; Tsvetanov and Earnhart, 2020). Specifically, the U.S. Department of Agriculture's (USDA) Conservation Reserve Enhancement Program (CREP) partners with state and local agencies to tailor conservation initiatives to local needs. In the HPA region, CREPs in Colorado, Kansas, and Nebraska address groundwater depletion concerns by paying groundwater users to retire their water rights and/or take irrigated lands out of production. Rosenberg (2020) and Manning et al. (2020) analyze the Kansas CREP and find that federal and state investments in well retirement yield groundwater conservation benefits but that the benefits likely do not outweigh the budgetary costs associated with the investments.

Figure 1 maps the extent of the HPA and highlights the study area regions where the social comparison interventions were introduced. Specifically, we focus on regions of the Colorado and Kansas HPA under the jurisdiction of Groundwater Management Districts

(GWMD). The remaining text in this section focuses on the institutional, climatic, and hydrological contexts of the Colorado and Kansas field experiment locations.

[Figure 1 about here.]

2.1 Background: Colorado

The Colorado field experiment focuses on groundwater irrigators in the Republican River Basin (RRB), which is a hydrologically connected sub-basin of the HPA. Corn, wheat, alfalfa, and dry beans are the most commonly irrigated crops in the RRB accounting for more than 92% of total irrigated land in the region during the 2021 growing season (CODNR, 2021). The Colorado Groundwater Commission (CGWC) is responsible for adjudicating groundwater rights and issuing use permits in the RRB. Use permits designate an annual volumetric appropriation although in practice these appropriations are not binding.[2] No formal well drilling moratoria or well spacing rules exist in the RRB. However, an informal well drilling moratoria does exist to some extent as the CGWC has not issued a new well permit since 2003. Seven local GWMDs manage the aquifer resources of the RRB of Colorado (see figure 1). Groundwater pumping in the RRB is unpriced and irrigators only pay the marginal cost associated with energy used to extract groundwater. There are no pumping quantity restrictions in the RRB outside of the volumetric appropriations associated with each well permit.

2.2 Background: Kansas

The Kansas field experiment concentrates on groundwater irrigators within the state's GWMDs (sometimes referred to as GMDs in Kansas). The HPA underlies all of the

[2]The volumetric groundwater appropriations associated with Colorado well permits were assigned when the permits were first issued. Changes in irrigation technology (e.g., the advent of center pivot irrigation) have rendered nearly all volumetric appropriations non-binding. For example, between 2011 and 2020, average well-level volumetric appropriations exceeded annual observed pumping by between 120 and 230 acre-feet depending on the year.

GWMDs that are located in western and central Kansas (see figure 1). The Kansas field experiment focuses on groundwater irrigators in 4 of the state's 5 GWMDs, which were created after the passage of legislation in 1972 (Kansas Statutes Annotated 82a-1027). Ongoing legal issues in GWMD 5 precluded including irrigators from that GWMD in the field experiment. The most commonly irrigated crops in Kansas are corn, wheat, alfalfa, soybeans, and sorghum, which jointly account for more than 88% of all irrigated acreage in the study area in 2021. To manage groundwater resources, Kansas's GWMDs initially used moratoria on well drilling and regulations on well spacing (Edwards, 2016). The passage of legislation in 2012 extended the statutory authorities of Kansas's GWMDs to allow for the creation of Local Enhancement Management Areas (LEMA). LEMAs allow GWMDs to develop their own more localized groundwater conservation plans, which are enforced by the state. There were three LEMAs in place during our intervention in Kansas. The Sheridan 6 LEMA began in 2013 and mandates significant reductions in water use, but is only the size of about 10% of a county. The GWMD 4 LEMA began in 2018 and covers a large area but allocations in the LEMA are often greater than historical use (Perez-Quesada and Hendricks, 2021). The Wichita County LEMA began in 2021 and imposes roughly a 25% reduction from historical use.

3 Experimental Design

In this section, we begin with a brief overview of the features of the experimental design that are common to the RCTs in both states. We then describe features that are specific to each state.

The target population for the experiment were owners or operators of groundwater wells that were permitted for irrigation and located in Colorado or Kansas. Multiple wells can be owned or operated by one individual/entity. To minimize inter-well spillovers (a.k.a., contamination), the assignment to treatment and control groups occurred at the well owner/operator level, not the well level (i.e., clustering the randomization at the individual/entity who controlled the wells reduced the potential for the treated wells to influence outcomes among the control wells). To associate wells with well owners/operators and their addresses, we used databases maintained by Colorado and Kansas. In some cases, multiple well owners/operators shared the same physical address. In these cases, we merged all well owners/operators sharing an address unit before assignment to treatment or control groups. Well owner/operator assignment to treatment and control groups is static over time (i.e., assignment to the treatment and control groups remains constant over the course of the experiment).

Prior to the beginning of the growing season, well owners/operators in the treatment group received a mailer with social comparison information. Control group well owners/operators received no mailer. The treatment mailers provided the recipient with well-level comparison information. Some producers owned/operated more than three wells, but the size constraints of the mailer led us to limit the well-level comparison information to the three wells that utilized the most water in the previous year. Figure 2 presents the distribution of well ownership patterns for Colorado and Kansas. Approximately 81% and 70% of Colorado and Kansas well owners/operators, respectively, own/operate 3 wells or less. Our primary analysis of the intervention's effect focuses only on wells that received comparison information or would have received comparison information if their owner/operator were in the treatment group. However, we also explore whether treatment affected those wells owned by treated owners/operators that did not receive comparison information.

[Figure 2 about here.]

Social comparison interventions typically combine a descriptive and injunctive norm (Ferraro et al., 2011; Jessoe et al., 2021b). To provide the injunctive norm, the treatment group mailers included the following statement:

Efficient water use is essential to managing our shared water resources.

Prior studies report that messages with injunctive norms alone nudge individual decision making toward pro-social behavior, but the effect is larger and persists longer when the injunctive norm is combined with a descriptive norm (Schultz et al., 2007; Ferraro et al., 2011). To provide a descriptive norm, the mailer graphically illustrated how the water consumption of the recipient's wells compared to water consumption by all wells in the comparison group. To provide this illustration, the mailer contrasted the water use of the recipient's wells in the previous year to the water use of the well at the 50^{th} and 20^{th} percentiles in the comparison group. Using text, rather than graphics, the mailer also reported the exact percentile of the well using the following language:

Comparing your [year] water use to other wells in the [comparison group], your well(s) recorded use higher than X% of wells.

Figure 3 presents examples of both the Colorado and Kansas mailers. The Colorado mailer was sent by the Department of Agricultural and Resource Economics at Colorado State University (CSU) and branded with CSU insignia. The Kansas mailer was sent by the Department of Agricultural Economics at Kansas State University (KSU) and branded with KSU insignia. The experimental protocols for both experiments were submitted to the universities' relevant Institutional Review Boards (IRB). At both universities, the experiments were deemed 'not human subject research' because no new data would be collected in the experiments.

[Figure 3 about here.]

To randomly assign well owners/operators to treatment and control groups, we blocked on the comparison group and past water use. The choice of which variables to use for the block randomization procedure was determined by modeling which variables best predict past observed water use (see registered pre-analysis plan for more information). To measure past water use in Kansas, we used the state's Water Information Management and Analysis System (WIMAS). To measure past water use in Colorado, we used data reported by the state's Department of Natural Resources. Both data systems record annual water use and irrigated acreage, but the acreage data in Colorado are not available until at least a year after a given growing season. To block randomize on past water use, we created a dummy variable, $Hi-Use_i^{own/op}$, which compares an individual owner/operator's past water use to water use in their comparison group. Specifically, $Hi-Use_i^{own/op}$ equals one if the well owner/operator's pre-experiment annual average water use across all their wells exceeds their comparison group's pre-experiment annual median water use and zero otherwise.

Power analyses aided the design of both the Colorado and Kansas interventions such that both experiments had the necessary statistical power to detect a 3% or greater reduction in water use. Pre-experiment power analyses were conducted with covariate control and assumed statistical power of 0.8 and α of 0.05 (see registered pre-analysis plan 'Supporting Documents and Materials' for more information). Table 1 presents summary statistics related to water use and growing season weather conditions for Colorado and Kansas treatment and control group wells between 2011 and 2021.

[Table 1 about here.]

Spatial patterns of well ownership/operation and data availability/quality issues created several experimental design challenges. First, a sizable proportion of irrigated agricultural land and associated wells in Colorado and Kansas are rented. Ideally, comparison mailers would be sent to the agricultural producer making water use decisions during the experiment. Identifying the well operator is possible in Kansas but not in Colorado. In Kansas, we define the well operator as the individual/entity that submitted the water use report, which is recorded in WIMAS. In contrast, Colorado does not regularly collect information on irrigated land rental patterns and the data linking operators to irrigated land/wells is dated. Given the likelihood that irrigated land rental patterns evolve over time while ownership remains relatively constant, we opted to send comparison mailers in Colorado to well owners, which include both operators and non-operator landlords (Bawa and Callahan, 2021). The second design challenge arises because some well owners/operators in both states had wells in multiple comparison groups. For these well owners/operators, their comparison group in the social comparison intervention was the GWMD/county within which the majority of their wells were located. The third design challenge arises because of errors or spelling variations in mailing addresses, which can prevent treated units from being exposed to treatment (i.e., non-compliance). Some addresses associated with well owners/operators were faulty or outdated, resulting in treatment mailers being returned by the postal service. Moreover, variations in abbreviation rules within the address data resulted in some well owners/operators receiving multiple mailings. In other words, a common address with spelling variations was viewed as two addresses.[3] This error, in some cases, led to a well owner/operator showing up more than once in the treated group (i.e., multiple mailers) or in the control group. In other cases, this error led to a well owner/operator being assigned to both treatment and control groups. We address this issue by dropping all wells associated with owners/operators assigned to both treatment and control groups from our subsequent analysis.

3.1 Experimental Design: Colorado

We identified 973 eligible well owners that collectively owned 2,658 wells in seven GWMDs. Comparison groups for the Colorado intervention were based on GWMD boundaries. The block randomization protocol resulted in 14 mutually-exclusive blocking groups (7

[3] A common example involves abbreviating or spelling out County Road (i.e., C.R. vs. County Road). For example, the address database contains two entities Big Farm LLC. and Bigger Farm LLC. with associated addresses of 123 County Road 4, Yuma, CO and 123 C.R. 4, Yuma, CO. Since C.R. is an abbreviation of 'County Road,' these two entities share a common address and should have been classified as a single treatment unit. Instead, when assigning well owners/operators to treatment and control groups these entities were sometimes treated as separate. Ideally, these duplicates would have all been identified prior to randomization to treatment and control groups. However, we discovered these duplicated addresses after assignment to treatment and control.

GWMDs X 2 pre-treatment water use groups).

Before randomizing well owners into treatment and control groups, we dropped outlier wells based on criteria described in the experiment's preregistration. Specifically, we dropped any wells that used less than their comparison groups 5^{th} percentile in the growing season before the experiment began (2018). We excluded these lower water use wells from the experiment given the likelihood that these wells were used for different purposes (i.e., watering livestock) than the targeted experiment population (i.e., irrigation). This issue is prevalent in Colorado where well permits can designate multiple uses (i.e., irrigation and livestock watering). Winnowing Colorado well observations based on these criteria resulted in 967 well owners associated with 2,529 wells. The number of eligible well owners decreased because in several cases all of the wells associated with a given well owner met the 5^{th} percentile outlier criteria. Implementing the randomization procedure among the 967 eligible Colorado well owners resulted in a control group comprised of 484 well owners associated with 1271 wells and a treatment group comprised of 483 well owners associated with 1258 wells.

Prior to the 2019, 2020, and 2021 growing seasons, Colorado well owners in the treatment group received mailers that provided information comparing their well-level water use in acre-feet in the previous growing season to other wells in their comparison group.[4] Comparisons were in total water used rather than water used per acre because the irrigated acreage data from the previous growing season did not appear in the state database until after the current growing season started. In the mailer, each individual well was identified by both its water diversion identifier (WDID) and its well permit number.

We limited the data used to evaluate the impact of the Colorado intervention according to attrition rules outlined in the experiment's pre-analysis plan. Specifically, we drop

[4]An error while printing the mailers sent to Colorado well owners prior to the 2019 growing season resulted in mismatched addresses and water use data. A corrected mailing was subsequently sent before the 2019 growing season started. The correct mailing provided the correct information as well as a message stating "Due to a printing error, the information in the mailer you received previously may have been incorrect. We apologize for the error and are now sending a correction."

any well observations that cease to pump or report water use after treatment randomization. Filtering the Colorado data based on these attrition criteria results in 957 well owners associated 2,497 wells remaining in the sample which translates into approximately a 1% attrition rate among both well owners and wells. Dropping these wells implicitly assumes these observations are missing independent of potential outcomes (MIPO). We further constrain the Colorado data by dropping any wells associated with a well owner that was incorrectly placed in both treatment and control groups due to address abbreviation errors. This results in a total of 863 well owners associated with 2,203 wells remaining in the Colorado data. Among these remaining well owners and wells, 431 well owners associated with 1,105 wells make up the control group while the treatment group consists of 432 well owners and 1,098 wells. Of these 2,203 total remaining wells, 1,614 wells were eligible to receive comparison information based on predefined rules limiting comparison information to a well owner's top three wells. The treatment compliance rate i.e., the share of treatment group mailers successfully delivered, for the Colorado experiment was 88%.

3.2 Experimental Design: Kansas

We identified 2,828 eligible well operators in Kansas that collectively own 10,243 wells.[5] Given that Kansas's GWMDs are relatively large and heterogeneous in terms of precipitation and groundwater availability, we opted to use counties to formulate the comparison groups. Several counties in Kansas contain less than 50 wells. To avoid disclosing personal information in the comparison mailers, we join these counties with fewer than 50 wells to adjacent counties to form a comparison group of 2 counties. In total this results in 4 of the 25 comparison group counties being aggregations of two counties.

[5] Kansas does not collect data on water use strictly at the well-level. Rather, water use data is collected at the water right level which can in some cases be an aggregation of several wells. For this experiment, we consider a 'well' to be a unique combination of water right number, water right identifier, and person identifier. When multiple water diversion records are associated with the same 'well' these diversion records are aggregated.

The Kansas experiment's block randomization process resulted in 50 mutually-exclusive blocking groups (25 counties X 2 pre-experiment water use groups).

Before randomizing well operators into treatment and control groups, we dropped outliers wells based on criteria described in the experiment's preregistration. Specifically, we dropped all wells that reported groundwater use (in acre-feet per acre irrigated) greater than the 99^{th} percentile of all Kansas wells during the pre-experiment growing season (2019). Additionally, we excluded all wells that reported irrigating more than 640 acres during the pre-experiment growing season. We excluded these high use wells from the Kansas experiment given the likelihood of inaccurate irrigated acreage or water use reporting. This is important for the Kansas experiment as in 2020 the state began using a new web-based system for water use reporting which allowed for more timely reporting of water use but may have introduced additional data entry errors. Winnowing Kansas wells observations based on these acreage and water use criteria resulted in 2,767 well operators associated with 9,605 wells. The number of eligible well operators decreased because in some cases all the wells associated with a given well operator failed to meet our predefined criteria. Implementing the randomization procedure among these 2,767 well operators resulted in a control group comprised of 1,381 well operators associated with 4,766 wells and a treatment group made up of 1,386 well operators associated with 4,839 wells.

Prior to the 2020 and 2021 growing seasons, Kansas's treatment group well operators received mailers providing recipients with information comparing their well-level water use in acre-feet per acre irrigated in the previous growing season to other well's within their comparison group. Individual well records were identified in the mailers using water right numbers (WRN) and the public land survey system section, township, and range of the parcel of land where the well is located.

We limited the data used to analyze the effect of the Kansas intervention according to the attrition rules outlined in the experiment's pre-registration. For the Kansas experiment, we define attrition as either a well ceasing to pump or report water use after treatment randomization or the individual/entity reporting a well's water use changing (i.e., a change in the well's ownership or rental status) after treatment randomization. Filtering Kansas well observations based on these attrition criteria results in 2,596 well operators associated with 8,863 wells which translates into approximately 6% and 8% attrition rates among well operators and wells, respectively. Dropping these wells implicitly assumes these observations are missing independent of potential outcomes (MIPO). We further constrain the Kansas data by dropping any wells associated with a well operator that was incorrectly placed in both treatment and control groups due to address abbreviation errors. This results in a total of 2,378 well operators associated with 7,733 wells remaining in the Kansas data. Among these remaining well operators and wells, 1,183 well operators associated with 3,839 wells make up the control group while the treatment group consists of 1,195 well operators and 3,894 wells. Of these 7,733 total remaining wells, 4,605 wells were eligible to receive comparison information based on predefined rules limiting comparison information to a well operator's top three wells. The treatment compliance rate i.e., the share of treatment group mailers successfully delivered, for the Kansas experiment was 96%.

4 Empirical Model

To estimate the treatment effect of the social comparison mailer on groundwater pumping behavior, we leverage a rich panel data set of well-level annual water use from 2011 to 2021. Specifically, we estimate the following pre-registered econometric model:

$$log(w_{i,j,t}) = \beta * Treatment_{j,t} + \alpha_i + \gamma * X_{i,j,t} + \varepsilon_{i,t}. \tag{1}$$

The dependent variable, $log(w_{i,j,t})$, is log transformed annual total water use by the i^{th} well associated with the j^{th} owner/operator in year t. $Treatment_{j,t}$ is a dummy variable which equals one if the j^{th} owner/operator of the i^{th} well received a mailer in year t communicating comparison information for the i^{th} well. β is the primary coefficient of interest which estimates the average treatment effect of the intervention. As the comparison intervention provided information for at most three wells, our primary analysis excludes treatment group wells that did not receive comparison information and control group wells that would not have received comparison information if they were assigned to the treatment group. In later analysis, we evaluate how treatment affects pumping behavior among treatment group wells that did not receive comparison information. α_i is a well-level random effect which accounts for well-level time invariant unobservables (e.g., soil type, owner/operator management ability, etc.). $X_{i,j,t}$ is a matrix of control variables and γ is the corresponding vector of coefficient estimates. Control variables include cumulative growing season precipitation, average growing season temperature, and the experiment's randomization blocking variables (comparison group indicator variables and $Hi-Use^{own/op}$). All control variables are time-varying with the exception of the comparison group indicator variables. $\varepsilon_{i,t}$ is an idiosyncratic error term. We control for well-level unobserved characteristics with a random-effects estimator rather than a fixed-effects estimator because the random-effects estimator is more efficient (Cameron and Trivedi, 2005; Allison, 2009; Abadie et al., 2022). The fixed-effects estimator would be preferred when there is potential bias from unobserved, time-invariant factors that create a correlation between the error term and the treatment variable, but that bias is not present in our design because the treatment was randomized.

Since treatment was randomized at the well owner/operator level (i.e., clusters of wells), the estimated standard errors for the coefficients are clustered at the well owner/operator level (Cameron and Miller, 2015; Imbens and Kolesar, 2016). For estimation, we used Bell and McCaffrey's (2002) leave-one-cluster-out jackknife variance estimator, implemented using the R package *clubSandwich* with Satterthwaite degrees of freedom (Satterthwaite, 1946; Pustejovsky, 2021). All empirical models were estimated using the R package *PLM*

(Croissant and Millo, 2008; R Core Team, 2019).

We extend the model specification outlined in equation 1 to explore conditional average treatment effects by including a suite of interaction variables. Specifically, to explore heterogeneity in treatment effects, we look at the moderating effect of two pre-registered covariates, $Hi-Use_{j,t}^{own/op}$ and $Hi-Use_{i,j,t}^{well}$. $Hi-Use_{i,j,t}^{well}$ is the well-level corollary of $Hi-Use_{j,t}^{own/op}$, that is $Hi-Use_{i,j,t}^{well}$ equals one in experiment year t if the i^{th} well used more water in the pre-experiment time period than their comparison group's median water use. Terms interacting these covariates with the treatment variable estimate treatment effects conditional on pre-experiment water use patterns. Given the high degree of correlation between $Hi-Use_{j,t}^{own/op}$ and $Hi-Use_{i,j,t}^{well}$, particularly for owners/operators associated with a single well, we separately estimate the moderating impact of pre-experiment owner/operator- and well-level water use patterns in two separate model specifications. Past research in the context of residential water use finds that social comparison interventions are most effective among high water use households (Ferraro and Price, 2013; Ferraro and Miranda, 2013; Brent et al., 2020). However, little is known about the prevalence of social comparison treatment effect heterogeneity in the context of agricultural water use, particularly how pre-experiment water use affects responsiveness to treatment (Chabé-Ferret et al., 2019).

In the absence of non-compliance, Equation 1 generates an unbiased estimate of the average treatment effect (ATE). Yet, as noted above, we have one-way, non-compliance in our experiment from returned mailers. We could estimate the treatment effect "as assigned" (i.e., the intent to treat effect, ITT) (Gupta, 2011), but we are more interested in the effect of treatment than the effect of treatment assignment. To estimate the effect of the social comparisons on those who received the mailer (i.e., the average treatment effect on the treated, ATT), we employ an instrumental variable (IV) approach (Angrist et al., 1996). The IV approach addresses treatment non-compliance by using the randomized treatment assignment as an instrument for treatment receipt, which allows one to estimate the ATT when non-compliance is only in the treated group.[6]

5 Experimental Results

Table 2 displays results estimating the treatment effects of the social comparison intervention. Specifically, column (1) presents the ATT of the intervention estimated by the IV model described in equation 1. Columns (2) and (3) present ATTs conditioning on owner/operator- and well-level pre-experiment water use patterns, respectively. For visual simplicity, table 2 only presents estimated coefficients for treatment and moderating covariates interacted with treatment. See Appendix A for the full suite of control and blocking variable coefficient estimates. The point estimate for the average treatment effect suggests that receiving a social comparison mailer reduces subsequent water use by approximately 4% .[7] Applying this estimated change in water use to the median control group well, which represents counterfactual water use absent the intervention, suggests that the mailer reduces subsequent water use by approximately 6.5 acre-feet. Aggregating estimated water use changes across control group wells indicates that the intervention decreased aggregate annual groundwater use by more than 27,000 acre-feet. This reduction in annual aggregate groundwater pumping is roughly equivalent to retiring 160 wells

[6]The validity of the IV approach for addressing treatment non-compliance rests on 5 assumptions explained in detail in Angrist et al. (1996). The assumptions are as follows: 1) the Stable Unit Treatment Value Assumption (SUTVA) posits that 'each unit has only one potential outcome per treatment value' (i.e., there is no interference between treatment units and there are not multiple versions of treatment), 2) random assignment to treatment, 3) the exclusion restrictions (i.e., treatment assignment impacts the outcome only through treatment receipt), 4) non-zero causal effect of treatment assignment on treatment receipt, and 5) monotonicity, which in our scenario presupposes that there are no cases where individuals always receive the opposite of what their treatment assignment indicates e.g., getting a comparison mailer when in the control group. We argue that both the Colorado and Kansas experiments meet the criteria delineated by these assumptions. Specifically, the SUTVA holds in our setting as treatment compliance relates to address reliability which would likely not affect other individual's receipt of treatment. Given that treatment assignment was random (see section 3) and impacts water use only through the receiving the comparison mailer which is causally linked to treatment group assignment our setting meets assumptions 2) through 4). Finally, both experiments meet the monotonicity assumption as there are no cases where control group well owners/operators consistently receive a comparison mailer.

[7]Throughout this section we use methods outlined in Halvorsen et al. (1980) to translate model point estimates into percentages.

with median levels of groundwater use for a year (see appendix B for more information).

[Table 2 about here.]

Columns (2) and (3) of table 2 present conditional average treatment effects that are specific to differing subgroups of wells based on pre-experiment water use (i.e., the covariates $Hi-Use^{well}$ and $Hi-Use^{own/op}$). Because assignment to $Hi-Use^{well}$ and $Hi-Use^{own/op}$ was not randomized but instead based on pre-experiment pumping behavior these coefficient estimates have a descriptive rather than causal interpretation (Emsley et al., 2010). Our analysis of conditional average treatment effects finds that treatment effects attenuate according to patterns of pre-experiment water use with the largest percentage-wise effects concentrating among wells and well owners/operators that used the least water in the pre-experiment time period. Specifically, we find that wells owned/operated by individuals/entities with lower than average pre-experiment water use ($Hi-Use^{own/op}=0$) and wells with lower than average pre-experiment water use ($Hi-Use^{well}=0$) exhibit a larger reduction in groundwater use of about 6% in response to treatment. Conversely, we find that diminished treatment effects among wells owned/operated by individuals/entities with higher than average pre-experiment water use ($Hi-Use^{own/op}=0$). These wells reduce groundwater use by approximately 1% to 2% in response to treatment. We find similar attenuated treatment effects among wells with higher than average pre-experiment water use. Together, these results highlight the importance of pre-experiment resource use patterns, which may be a proxy for resource conservation preferences, in determining the net impact of comparison interventions aiming to alter resource use decisions (Cherry et al., 2017).

Section C of the appendix tests the robustness of our primary results to differing model specifications and assumptions. In appendix C.1, we estimate equation 1 specifying α_i as a fixed effect and find qualitatively similar results to the random effect specification. In appendix C.2, we separately estimate equation 1 for each state's experiment and find qualitatively similar but less precise results. We also test the robustness of our results to issues arising from multiple mailers/addresses (see section 3). Appendix C.3 estimates experiment treatment effects with 1) a subset of data that drops well observations associated well owners/operators that were erroneously placed in treatment or control groups more than once and 2) by incorporating well owners/operators that were erroneously placed in both treatment and control groups and addressing this issue using empirical techniques outlined in Angrist et al. (1996). In both cases, we find qualitatively similar results to those presented in table 2. In appendix C.4, we take a cross-sectional modeling approach outlined in an early version of the experiment's pre-analysis plan to estimate treatment effects. Results are qualitatively similar although less precise than those generated using a panel data approach. We also test the robustness of our results to the IV approach leveraged to address treatment non-compliance. Appendix C.5 estimates an 'intent to treat' model and finds qualitatively similar results to the IV approach. Finally, in appendix C.6 we estimate a suite of models estimating temporal heterogeneity in treatment effects and find no evidence of treatment effect attenuation over time.

5.1 Within-irrigator Spillover Effects

Our primary analysis of the intervention's effect focuses on wells that either received comparison information or would have if assigned to the treatment group. Comparison information was not provided to all wells owned/operated by individuals/entities assigned to the treatment group as size constraints in the comparison mailers precluded providing information for more than three wells (see section 3 for more information). Here we empirically test for any spillover effects of the intervention. Namely, we test how a well owner's/operator's inclusion in the treatment group affects their pumping behavior among their non-treated wells (i.e., those wells they own/operate that did not receive any comparison information). To test for these spillover effects we estimate the previously described econometric models focusing on treatment group wells that received no comparison information and control group wells that would not have received comparison information if their owner/operator had been assigned to the treatment group.

[Table 3 about here.]

Table 3 presents results of the models estimating the within-irrigator spillover effects of the intervention. For simplicity, table 3 focuses only on covariates related to treatment. A table with the full set of regressions can be found in appendix A. Results suggest that the comparison intervention generated significant spillovers among non-treatment wells owned/operated by individuals in the treatment group. Specifically, wells owned/operated by treatment group individuals/entities that did not receive any comparison information reduced groundwater pumping by approximately 4.3%, a result that qualitatively aligns with treatment effects estimated among wells that received comparison information.

Columns (2) and (3) of table 3 present results for the specifications estimating the moderating influence of pre-experiment water use patterns on treatment effects. Results largely align with the conditional average treatment effects presented in table 2 with treatment generating the largest impact among those wells and owners/operators already using below average quantities of groundwater. However, within-irrigator spillover model results are generally less precise than those estimated for wells actually receiving comparison information.

6 Conclusion

This paper builds on the existing economics literature utilizing randomized control trials to understand the impacts of social comparison interventions. We present results from a novel social comparison field experiment implemented among groundwater irrigators in Colorado and Kansas. The intervention provided an informational mailer to a randomized group of agricultural groundwater users comparing their prior groundwater use to that of their neighbors. We find that the intervention induced an economically significant reduction in subsequent groundwater use. Our results demonstrate the efficacy of 'nudge' type interventions to help alleviate the common pool and public good issues prevalent among many environmental goods and natural resources.

Results of the Colorado and Kansas comparison interventions have important implications for the management of increasingly scarce groundwater resources. We find that receiving social comparison information reduces subsequent groundwater use by approximately 4%. Applying the estimated effect of the comparison mailer to counterfactual water use in the absence of treatment represented by control group water use suggests that the intervention reduced aggregate annual groundwater use by more than 27,000 acre-feet, which is roughly equivalent to retiring 160 wells for a year. The aggregate change in groundwater use induced by the comparison intervention reduces annual groundwater depletion by 1.65% and 1.03% in Colorado and Kansas, respectively (see appendix B for more information). These results demonstrate the efficacy of comparison based behavioral nudges in helping to alleviate groundwater depletion issues common in many of the most agriculturally productive aquifers (Scanlon et al., 2012; Bierkens and Wada, 2019).

To place our results within the context of other policy efforts aiming to address groundwater depletion concerns (e.g., well and irrigated land retirement programs), we calculate the per acre-foot conserved cost of the intervention. This analysis indicates per acre-foot conserved costs of the intervention of \$0.10 and \$0.09 for Colorado and Kansas, respectively (see Appendix D for more information). The estimated per acre-foot conserved cost of the comparison intervention is approximately $1/200^{th}$ the annual per acre-foot conserved cost of a well retirement program implemented in the study area (Tsvetanov and Earnhart, 2020). This result suggests that compared to other policy interventions, behavioral nudges may constitute a relatively cost effective means to conserve certain environmental goods and natural resources.

We also find that intervention treatment effects persist among treatment group wells that did not receive comparison information. The within-irrigator spillovers arising from the comparison field experiment demonstrate how interventions can alter behavior even among non-treated units belonging to treatment group individuals or firms. This insight contributes to an emerging literature on within treatment group spillovers generated by behavioral interventions (e.g., Carpenter and Lawler (2019) and Jessoe et al. (2021b)) and can inform future behavioral interventions where data availability or other limitations preclude fully treating all relevant units associated with treatment group entities.

This behavioral intervention and its results constitute an important contribution to the social comparison literature. Specifically, we provide novel evidence regarding the efficacy of comparison interventions implemented among firms rather than consumers. The majority of the literature focuses on altering consumer behavior through the provision of comparison information (Allcott, 2011; Ferraro and Price, 2013; Bernedo et al., 2014; Brent et al., 2015). Notable exceptions include Earnhart and Ferraro (2020), Wallander et al. (2017), and Chabé-Ferret et al. (2019), who test the efficacy of social comparison interventions on water treatment firm behavior in Kansas, U.S. agricultural producers' decisions to enroll in conservation programs, and surface water use among irrigated farmers in France, respectively. Our work most directly builds on Chabé-Ferret et al. (2019) who report results from a relatively under-powered experiment and find no statistically discernible effects of comparison messaging on agricultural water use leaving important questions unanswered regarding the efficacy of social comparisons in altering agricultural firms water use behavior. This paper addresses these questions using a sufficiently powered field experiment whose results demonstrate the efficacy of comparison interventions in altering the water use behavior of profit-seeking agricultural firms.

Our results also have important implications for the management of other environmental goods and natural resources. Experimental results demonstrate that social comparison messaging can influence the behavior of agricultural producers toward pro-social outcomes. This result is significant given the substantial role that agricultural producer decision-making has in determining the status of many natural resource stocks and the environment. Social comparison interventions could potentially be used to encourage the adoption of conservation practices (e.g., cover cropping, no-till, etc.) or discourage practices that generate negative environmental externalities (e.g., excessive fertilizer application). Future experimental research is needed investigating the efficacy of social comparison interventions in these diverse settings.

Finally, important questions remain regarding the persistence of treatment effects over time and whether agricultural producers habituate to repeated treatment. In the context of residential energy and water demand, past research has found treatment effects persist over time and consumers are slow to habituate to continued interventions (Ferraro et al., 2011; Allcott and Rogers, 2014). Our results from three years of interventions suggest that the treatment effects do not substantially attenuate amongst groundwater irrigators, but more research is needed to understand long-run responses. If treatment effects persist with repeated interventions, then sustained social comparison interventions may offer a cost-effective means to diminish groundwater overdraft and promote pro-social behavior. If not, then there may be opportunities to intermittently implement interventions through time to address groundwater depletion.

References

Abadie, A., Athey, S., Imbens, G. W., and Wooldridge, J. M. (2022). When should you adjust standard errors for clustering? *The Quarterly Journal of Economics*.

Allcott, H. (2011). Social norms and energy conservation. *Journal of Public Economics*, 95(9-10):1082–1095.

Allcott, H. and Mullainathan, S. (2010). Behavior and energy policy. *Science*, 327(5970):1204–1205.

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.

Allison, P. D. (2009). *Fixed effects regression models*. SAGE publications.

Andor, M. A. and Fels, K. M. (2018). Behavioral economics and energy conservation – a systematic review of non-price interventions and their causal effects. *Ecological Economics*, 148:178–210.

Angrist, J. D., Imbens, G. W., and Rubin, D. B. (1996). Identification of causal effects using instrumental variables. *Journal of the American statistical Association*, 91(434):444–455.

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.

Barrett, C. B. and Carter, M. R. (2020). Finding our balance? revisiting the randomization revolution in development economics ten years further on. *World development*, 127:104789.

Bawa, S. G. and Callahan, S. (2021). Absent landlords in agriculture–a statistical analysis. Technical report.

Bell, R. M. and McCaffrey, D. F. (2002). Bias reduction in standard errors for linear regression with multi-stage samples. *Survey Methodology*, 28(2):169–182.

Bernedo, M., Ferraro, P. J., and Price, M. (2014). The persistent impacts of norm-based messaging and their implications for water conservation. *Journal of Consumer Policy*, 37(3):437–452.

Beshears, J., Choi, J. J., Laibson, D., Madrian, B. C., and Milkman, K. L. (2015). The effect of providing peer information on retirement savings decisions. *The Journal of finance*, 70(3):1161–1201.

Bhanot, S. P. (2017). Rank and response: A field experiment on peer information and water use behavior. *Journal of Economic Psychology*, 62:155–172.

Bierkens, M. F. and Wada, Y. (2019). Non-renewable groundwater use and groundwater depletion: a review. *Environmental Research Letters*, 14(6):063002.

Brent, D. A., Cook, J. H., and Olsen, S. (2015). Social comparisons, household water use, and participation in utility conservation programs: Evidence from three randomized trials. *Journal of the Association of Environmental and Resource Economists*, 2(4):597–627.

Brent, D. A., Lott, C., Taylor, M., Cook, J., Rollins, K., and Stoddard, S. (2020). What causes heterogeneous responses to social comparison messages for water conservation? *Environmental and Resource Economics*, 77(3):503–537.

Cameron, A. C. and Miller, D. L. (2015). A practitioner's guide to cluster-robust inference. *Journal of human resources*, 50(2):317–372.

Cameron, A. C. and Trivedi, P. K. (2005). *Microeconometrics: methods and applications*. Cambridge university press.

Carlsson, F., Gravert, C., Johansson-Stenman, O., and Kurz, V. (2021). The use of green nudges as an environmental policy instrument. *Review of Environmental Economics and Policy*, 15(2):216–237.

Carlsson, F. and Johansson-Stenman, O. (2012). Behavioral economics and environmental policy. *Annual Review of Resource Economics*, 4(1):75–99.

Carpenter, C. S. and Lawler, E. C. (2019). Direct and spillover effects of middle school vaccination requirements. *American Economic Journal: Economic Policy*, 11(1):95–125.

Chabé-Ferret, S., Coent, P. L., Reynaud, A., Subervie, J., and Lepercq, D. (2019). Can we nudge farmers into saving water? evidence from a randomised experiment. *European Review of Agricultural Economics*, 46(3):393–416.

Cherry, T. L., Kallbekken, S., and Kroll, S. (2017). Accepting market failure: Cultural worldviews and the opposition to corrective environmental policies. *Journal of Environmental Economics and Management*, 85:193–204.

Cialdini, R. B., Reno, R. R., and Kallgren, C. A. (1990). A focus theory of normative conduct: Recycling the concept of norms to reduce littering in public places. *Journal of personality and social psychology*, 58(6):1015.

CODNR (2021). Colorado decision support systems, gis data by category, division 1 - republican irrigated lands. https://cdss.colorado.gov/gis-data/gis-data-by-category. created 1 Aug 2021.

Croissant, Y. and Millo, G. (2008). Panel data econometrics in R: The plm package. *Journal of Statistical Software*, 27(2):1–43.

Croson, R. and Treich, N. (2014). Behavioral environmental economics: promises and challenges. *Environmental and Resource Economics*, 58(3):335–351.

Deaton, A. (2010). Instruments, randomization, and learning about development. *Journal of economic literature*, 48(2):424–55.

Deines, J. M., Kendall, A. D., Butler, J. J., and Hyndman, D. W. (2019). Quantifying irrigation adaptation strategies in response to stakeholder-driven groundwater management in the us high plains aquifer. *Environmental Research Letters*, 14(4):044014.

Drysdale, K. M. and Hendricks, N. P. (2018). Adaptation to an irrigation water restriction imposed through local governance. *Journal of Environmental Economics and Management*, 91:150–165.

Earnhart, D. and Ferraro, P. J. (2020). The effect of peer comparisons on polluters: A randomized field experiment among wastewater dischargers. *Environmental and Resource Economics*.

Edwards, E. C. (2016). What lies beneath? aquifer heterogeneity and the economics of groundwater management. *Journal of the Association of Environmental and Resource Economists*, 3(2):453–491.

Edwards, E. C. and Guilfoos, T. (2020). The economics of groundwater governance institutions across the globe. *Applied Economic Perspectives and Policy*.

Emsley, R., Dunn, G., and White, I. R. (2010). Mediation and moderation of treatment effects in randomised controlled trials of complex interventions. *Statistical methods in medical research*, 19(3):237–270.

Famiglietti, J. S. (2014). The global groundwater crisis. *Nature Climate Change*, 4(11):945–948.

Feinerman, E. and Knapp, K. C. (1983). Benefits from groundwater management: magnitude, sensitivity, and distribution. *American Journal of Agricultural Economics*, 65(4):703–710.

Ferraro, P., Messer, K. D., and Wu, S. (2017). Applying behavioral insights to improve water security. *Choices*, 32(4):1–6.

Ferraro, P. J. and Agrawal, A. (2021). Sustaining the commons: Advancing understanding of common pool resource management special feature: Synthesizing evidence in sustainability science through harmonized experiments: Community monitoring in common pool resources. *Proceedings of the National Academy of Sciences of the United States of America*, 118(29).

Ferraro, P. J., Messer, K. D., Shukla, P., and Weigel, C. (2022). Behavioral biases among producers: Experimental evidence of anchoring in procurement auctions. *The Review of Economics and Statistics*, pages 1–40.

Ferraro, P. J. and Miranda, J. J. (2013). Heterogeneous treatment effects and mechanisms in information-based environmental policies: Evidence from a large-scale field experiment. *Resource and Energy Economics*, 35(3):356–379.

Ferraro, P. J., Miranda, J. J., and Price, M. K. (2011). The persistence of treatment effects with norm-based policy instruments: Evidence from a randomized environmental policy experiment. *American Economic Review*, 101(3):318–322.

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

Ferraro, P. J. and Shukla, P. (2020). Feature—is a replicability crisis on the horizon for environmental and resource economics? *Review of Environmental Economics and Policy*.

Ferraro, P. J. and Shukla, P. (2022). Credibility crisis in agricultural economics. *Applied Economic Perspectives and Policy*.

Festinger, L. (1954). A theory of social comparison processes. *Human Relations*, 7(2):117–140.

Frey, B. S. and Meier, S. (2004). Social comparisons and pro-social behavior: Testing" conditional cooperation" in a field experiment. *American Economic Review*, 94(5):1717–1722.

Gerber, A. S. and Rogers, T. (2009). Descriptive social norms and motivation to vote: Everybody's voting and so should you. *The Journal of Politics*, 71(1):178–191.

Goldstein, N. J., Cialdini, R. B., and Griskevicius, V. (2008). A room with a viewpoint: Using social norms to motivate environmental conservation in hotels. *Journal of Consumer Research*, 35(3):472–482.

Guilfoos, T., Khanna, N., and Peterson, J. M. (2016). Efficiency of viable groundwater management policies. *Land Economics*, 92(4):618–640.

Gupta, S. K. (2011). Intention-to-treat concept: a review. *Perspectives in clinical research*, 2(3):109.

Haacker, E. M. K., Kendall, A. D., and Hyndman, D. W. (2015). Water level declines in the high plains aquifer: Predevelopment to resource senescence. *Groundwater*, 54(2):231–242.

Hahn, R. and Metcalfe, R. (2016). The impact of behavioral science experiments on energy policy. *Economics of Energy & Environmental Policy*, 5(2):27–44.

Halvorsen, R., Palmquist, R., et al. (1980). The interpretation of dummy variables in semilogarithmic equations. *American economic review*, 70(3):474–475.

Hargreaves, G. H. and Samani, Z. A. (1985). Reference crop evapotranspiration from temperature. *Applied engineering in agriculture*, 1(2):96–99.

Hrozencik, R. A., Manning, D. T., Suter, J. F., and Goemans, C. (2021). Impacts of block-rate energy pricing on groundwater demand in irrigated agriculture. *American Journal of Agricultural Economics*, n/a(n/a).

Hrozencik, R. A., Manning, D. T., Suter, J. F., Goemans, C., and Bailey, R. T. (2017). The heterogeneous impacts of groundwater management policies in the republican river basin of colorado. *Water Resources Research*, 53(12):10757–10778.

Imbens, G. W. and Kolesar, M. (2016). Robust standard errors in small samples: Some practical advice. *Review of Economics and Statistics*, 98(4):701–712.

Ioannidis, J. P. A., Stanley, T. D., and Doucouliagos, H. (2017). The power of bias in economics research. *The Economic Journal*, 127(605):F236–F265.

Jessoe, K., Lade, G. E., Loge, F., and Spang, E. (2021a). Spillovers from behavioral interventions: Experimental evidence from water and energy use. *Journal of the Association of Environmental and Resource Economists*, 8(2):315–346.

Jessoe, K., Lade, G. E., Loge, F., and Spang, E. (2021b). Spillovers from behavioral interventions: Experimental evidence from water and energy use. *Journal of the Association of Environmental and Resource Economists*, 8(2):315–346.

Konikow, L. F. (2011). Contribution of global groundwater depletion since 1900 to sea-level rise. *Geophysical Research Letters*, 38(17).

Konikow, L. F. (2015). Long-term groundwater depletion in the united states. *Groundwater*, 53(1):2–9.

Konikow, L. F. and Kendy, E. (2005). Groundwater depletion: A global problem. *Hydrogeology Journal*, 13(1):317–320.

Lovelace, J. K., Nielsen, M. G., Read, A. L., Murphy, C. J., and Maupin, M. A. (2020). Estimated groundwater withdrawals from principal aquifers in the united states, 2015.

Manning, D. T., Rad, M. R., Suter, J. F., Goemans, C., Xiang, Z., and Bailey, R. (2020). Non-market valuation in integrated assessment modeling: The benefits of water right retirement. *Journal of Environmental Economics and Management*, 103:102341.

Monger, R. G., Suter, J. F., Manning, D. T., and Schneekloth, J. P. (2018). Retiring land to save water: Participation in colorado's republican river conservation reserve enhancement program. *Land Economics*, 94(1):36–51.

Montginoul, M., Rinaudo, J.-D., Brozović, N., and Donoso, G. (2016). Controlling groundwater exploitation through economic instruments: Current practices, challenges and innovative approaches. In *Integrated groundwater management*, pages 551–581. Springer, Cham.

Nieswiadomy, M. (1985). The demand for irrigation water in the high plains of texas, 1957–80. *American Journal of Agricultural Economics*, 67(3):619–626.

Palm-Forster, L. H., Ferraro, P. J., Janusch, N., Vossler, C. A., and Messer, K. D. (2019). Behavioral and experimental agri-environmental research: Methodological challenges, literature gaps, and recommendations. *Environmental and Resource Economics*, 73(3):719–742.

Perez-Quesada, G. and Hendricks, N. P. (2021). Lessons from local governance and collective action efforts to manage irrigation withdrawals in kansas. *Agricultural Water Management*, 247:106736.

Pfeiffer, L. and Lin, C.-Y. C. (2014). Does efficient irrigation technology lead to reduced groundwater extraction? empirical evidence. *Journal of Environmental Economics and Management*, 67(2):189–208.

PRISM (2021). Prism climate group, oregon state university. http://prism.oregonstate.edu. created 1 Aug 2021.

Pustejovsky, J. (2021). *clubSandwich: Cluster-Robust (Sandwich) Variance Estimators with Small-Sample Corrections*. R package version 0.5.3.

R Core Team (2019). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Reeves, R. (2012). Guideline, education, and peer comparison to reduce prescriptions of benzodiazepines and low-dose quetiapine in prison. *Journal of Correctional Health Care*, 18(1):45–52.

Riseng, C., Wiley, M., Black, R. W., and Munn, M. (2011). Impacts of agricultural land use on biological integrity: a causal analysis. *Ecological Applications*, 21(8):3128–3146.

Roe, B. E. and Just, D. R. (2009). Internal and external validity in economics research: Tradeoffs between experiments, field experiments, natural experiments, and field data. *American Journal of Agricultural Economics*, 91(5):1266–1271.

Rosenberg, A. B. (2020). Targeting of water rights retirement programs: Evidence from kansas. *American Journal of Agricultural Economics*, 102(5):1425–1447.

Sacarny, A., Yokum, D., Finkelstein, A., and Agrawal, S. (2016). Medicare letters to curb overprescribing of controlled substances had no detectable effect on providers. *Health Affairs*, 35(3):471–479.

Satterthwaite, F. E. (1946). An approximate distribution of estimates of variance components. *Biometrics bulletin*, 2(6):110–114.

Scanlon, B. R., Faunt, C. C., Longuevergne, L., Reedy, R. C., Alley, W. M., McGuire, V. L., and McMahon, P. B. (2012). Groundwater depletion and sustainability of irrigation in the US high plains and central valley. *Proceedings of the National Academy of Sciences*, 109(24):9320–9325.

Schultz, P. W., Nolan, J. M., Cialdini, R. B., Goldstein, N. J., and Griskevicius, V. (2007). The constructive, destructive, and reconstructive power of social norms. *Psychological Science*, 18(5):429–434.

Shang, J. and Croson, R. (2009). A field experiment in charitable contribution: The impact of social information on the voluntary provision of public goods. *The economic journal*, 119(540):1422–1439.

Slough, T., Rubenson, D., Levy, R., Alpizar Rodriguez, F., Bernedo del Carpio, M., Buntaine, M. T., Christensen, D., Cooperman, A., Eisenbarth, S., Ferraro, P. J., et al. (2021). Adoption of community monitoring improves common pool resource management across contexts. *Proceedings of the National Academy of Sciences*, 118(29):e2015367118.

Smith, S. M., Andersson, K., Cody, K. C., Cox, M., and Ficklin, D. (2017). Responding to a groundwater crisis: The effects of self-imposed economic incentives. *Journal of the Association of Environmental and Resource Economists*, 4(4):985–1023.

Steward, D. R. and Allen, A. J. (2016). Peak groundwater depletion in the high plains aquifer, projections from 1930 to 2110. *Agricultural Water Management*, 170:36–48.

Steward, D. R., Bruss, P. J., Yang, X., Staggenborg, S. A., Welch, S. M., and Apley, M. D. (2013). Tapping unsustainable groundwater stores for agricultural production in the high plains aquifer of kansas, projections to 2110. *Proceedings of the National Academy of Sciences*, 110(37):E3477–E3486.

Suárez, F. G., Fulginiti, L. E., and Perrin, R. K. (2018). What is the use value of irrigation water from the high plains aquifer? *American Journal of Agricultural Economics*, 101(2):455–466.

Torres, M. M. J. and Carlsson, F. (2018). Direct and spillover effects of a social information campaign on residential water-savings. *Journal of Environmental Economics and Management*, 92:222–243.

Tsioumani, E. (2019). The state of the world's biodiversity for food and agriculture: a call to action? *Environmental Policy and Law*, 49(2/3):110–112.

Tsvetanov, T. and Earnhart, D. (2020). The effectiveness of a water right retirement program at conserving water. *Land Economics*, 96(1):56–74.

Van Meter, K., Van Cappellen, P., and Basu, N. (2018). Legacy nitrogen may prevent achievement of water quality goals in the gulf of mexico. *Science*, 360(6387):427–430.

Wallander, S., Ferraro, P., and Higgins, N. (2017). Addressing participant inattention in federal programs: A field experiment with the conservation reserve program. *American Journal of Agricultural Economics*, 99(4):914–931.

Yoeli, E., Budescu, D. V., Carrico, A. R., Delmas, M. A., DeShazo, J., Ferraro, P. J., Forster, H. A., Kunreuther, H., Larrick, R. P., Lubell, M., et al. (2017). Behavioral science tools to strengthen energy & environmental policy. *Behavioral Science & Policy*, 3(1):68–79.

Figures

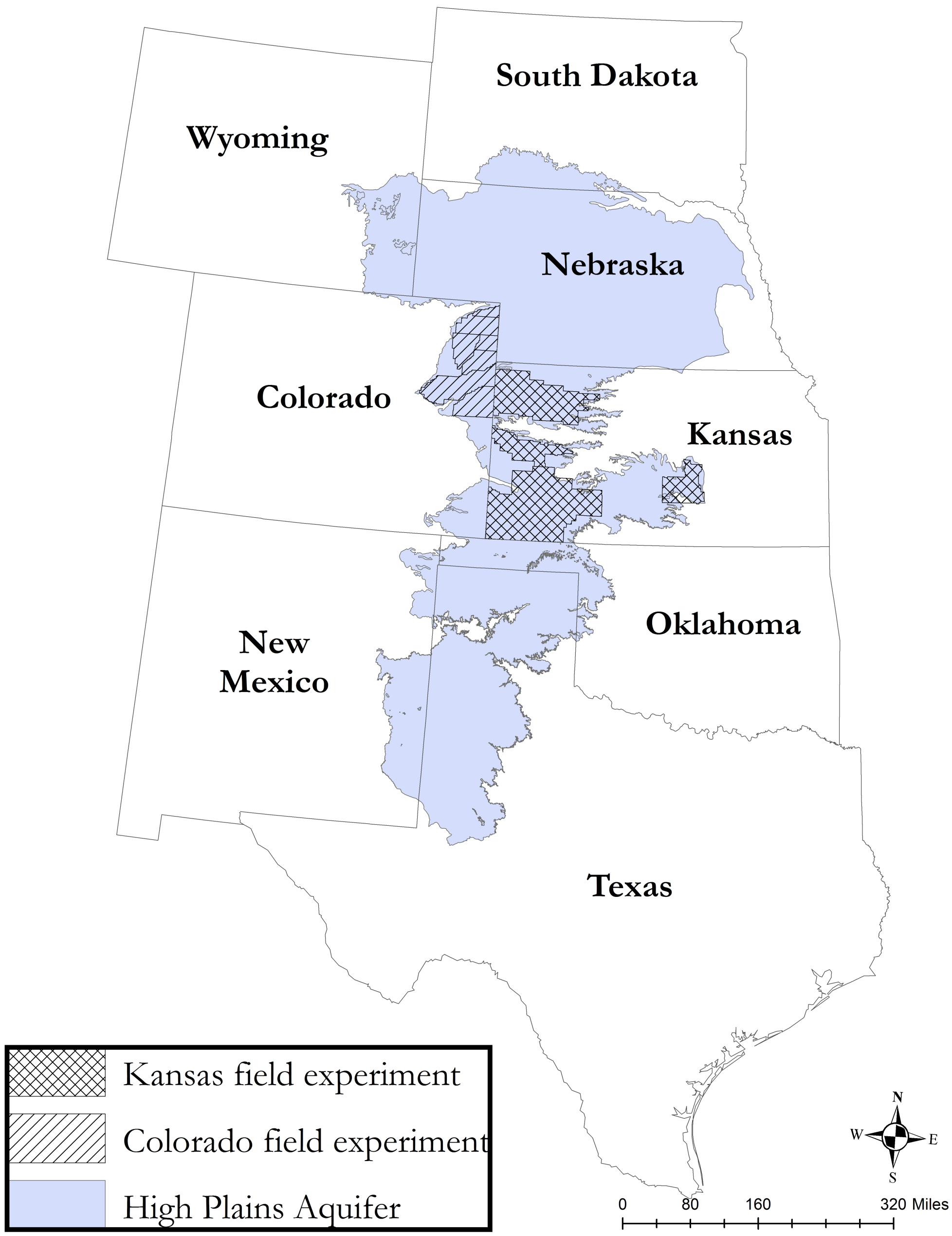


Figure 1: Colorado and Kansas study areas

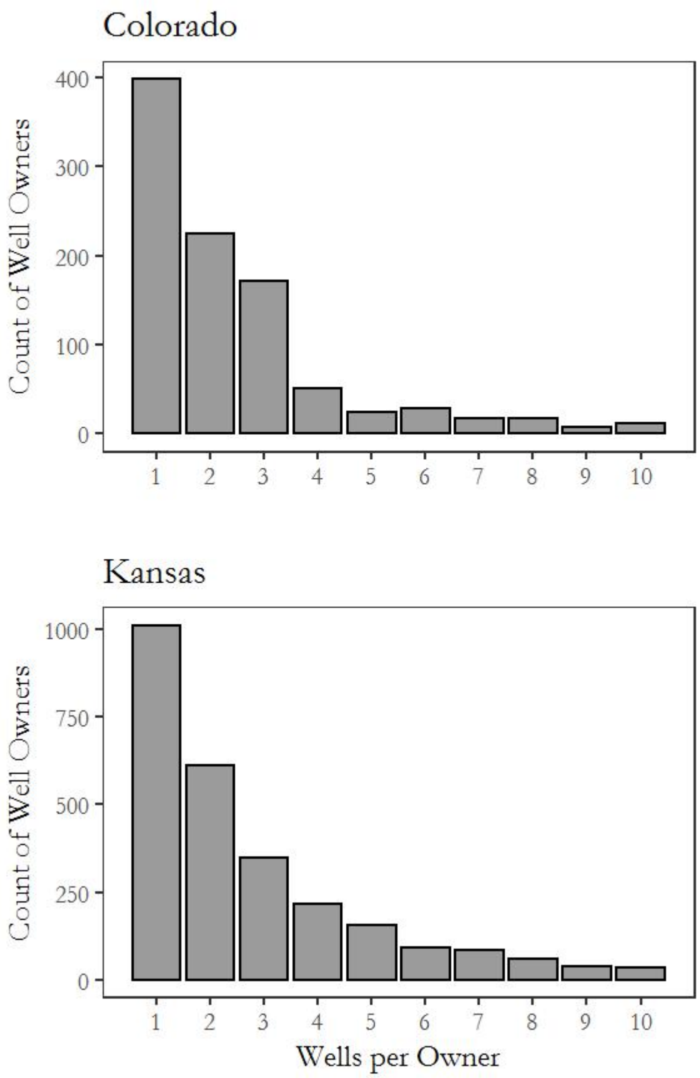
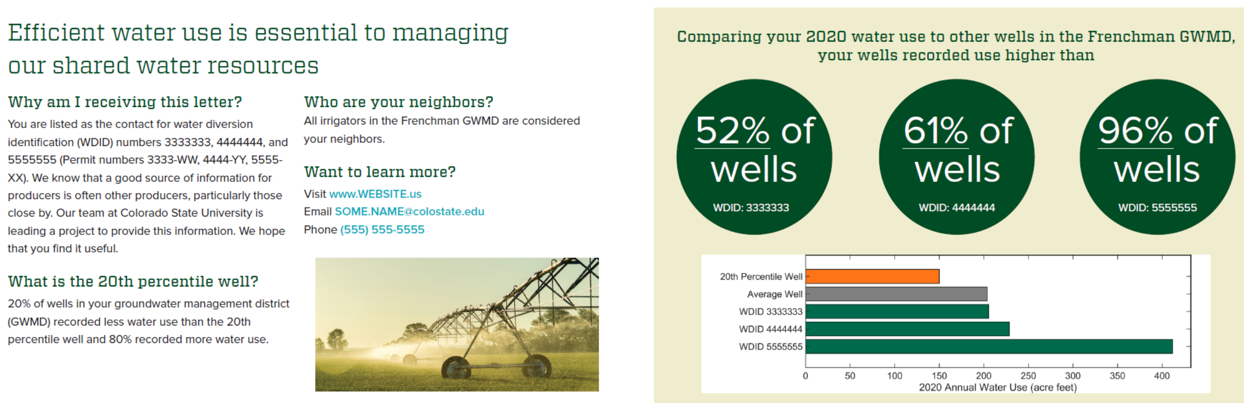
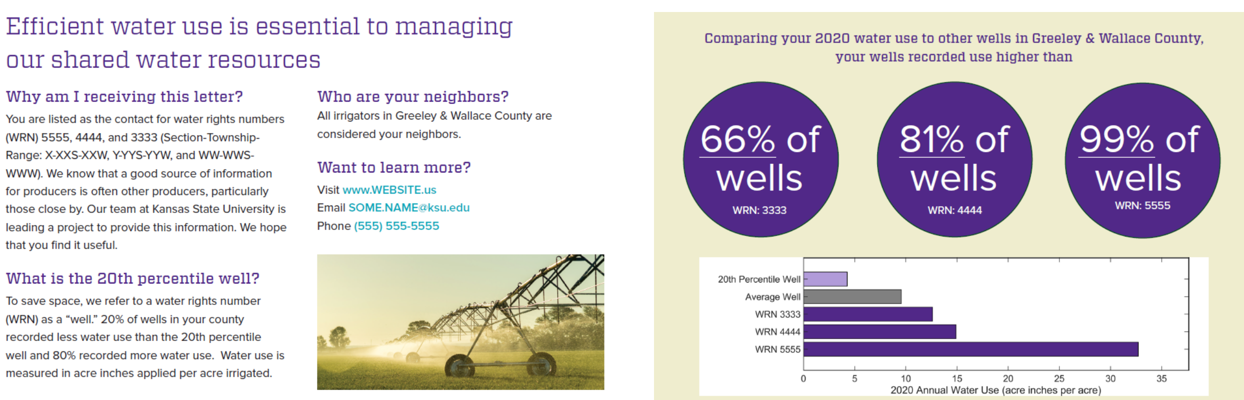


Figure 2: Irrigation wells per well owner
Note: For visual simplicity figure 2 does not display data for well owners/operators that have more than 10 wells as such concentrations of well ownership and operation are relatively rare. Approximately, 2% and 6% of well owners/operators in Colorado and Kansas, respectively, own or operate more than 11 irrigation wells.



Colorado mailer



Kansas mailer

Figure 3: Example social comparison mailers

Note: The above example mailers are based on actual comparison interventions sent to a Colorado and and Kansas treatment group well owner/operator in 2021. We have anonymized both mailers by replacing water diversion identification numbers (WDID), water rights numbers (WRN), permit numbers, and section-township-range indicators with fabricated entries.

Tables

Table 1: **Average Annual Groundwater Use and Weather Summary Statistics, 2011-2021**

Annual summary statistic	N	Mean	St. Dev.
Groundwater Use (acre feet)	66,292	189.09	126.78
Irrigated Acreage (acres)	59,857	162.43	112.80
Water User per Acre (acre inches per acre)	59,857	13.92	6.08
Temperature (° Fahrenheit)	66,292	75.65	2.53
Precipitation (inches)	66,292	9.22	3.65

The number of groundwater use observations is greater than irrigated acreage and water use per acre observations as acreage data are not available for Colorado between 2011 and 2014.
Temperature is the mean daily temperature across the growing season (June, July, and August).
Precipitation is the cumulative total of growing season (June, July, and August) rainfall.

Table 2: **Average Treatment Effect IV Model**

	Dependent variable:		
	Log(Annual Groundwater Use)		
	(1)	(2)	(3)
Treatment	-0.0414***	-0.0604***	-0.0576***
	(-0.0605,-0.0223)	(-0.0865,-0.0344)	(-0.0858,-0.0294)
Treatment X $Hi-Use^{own/op}$		0.0509**	
		(0.0166,0.0852)	
Treatment X $Hi-Use^{well}$			0.0400*
			(0.0056,0.0744)
Blocking Variables	Yes	Yes	Yes
Well Random Effect	Yes	Yes	Yes
Observations	66,292	66,292	66,292
R^2	0.4509	0.4510	0.4527
F Statistic	684,255.8000***	684,567.6000***	693,794.0000***
Note:	* p<0.05; ** p<0.01; *** p<0.001 Standard errors clustered at the well-owner/operator level. 95 percent confidence intervals in parentheses below coefficient estimates.		

Table 3: **Average Treatment Effect IV Within-irrigator Spillover Model**

	Dependent variable:		
	Log(Annual Groundwater Use)		
	(1)	(2)	(3)
Treatment	-0.0445**	-0.0575**	-0.0623**
	(-0.0745,-0.0144)	(-0.1011,-0.0139)	(-0.1018,-0.0227)
Treatment X $Hi-Use^{own/op}$		0.0318	
		(-0.0249,0.0886)	
Treatment X $Hi-Use^{well}$			0.0422
			(-0.0128,0.0973)
Blocking Variables	Yes	Yes	Yes
Well Random Effect	Yes	Yes	Yes
Observations	39,783	39,783	39,783
R^2	0.3767	0.3768	0.3773
F Statistic	385,887.1000***	386,176.3000***	387,894.8000***
Note:	* p<0.05; ** p<0.01; *** p<0.001 Standard errors clustered at the well-owner/operator level. 95 percent confidence intervals in parentheses below coefficient estimates.		

Appendices

A Full Empirical Modeling Results

For visual simplicity, the main text of this paper does not present results for the full set of covariates used to estimated equation 1. Here we present and discuss the full set of empirical results for the interested reader. Specifically, tables A.1 and A.2 displays full model results for the suite of specifications presented in tables 2 and 3 of the main text, respectively. Note that while the full models include comparison group indicator variables, tables A.1 and A.2 do not present these coefficient estimates as there are more than 30 individual comparison groups.

The estimated sign of the impact of the experiment blocking variable $Hi-Use^{own/op}$ on water use is mixed across model specifications and relatively imprecise. Additionally, the coefficient estimate associated with $Hi-Use^{well}$ suggests that wells with high pre-experiment water use reduced their water use in subsequent years (see column (3) of tables A.1 and A.2). However, this coefficient is only precisely estimated for the model analyzing treatment effects among treated wells. Finally, the impacts of temperature and precipitation on annual groundwater pumping follow intuition. There is a positive and significant relationship between temperature and groundwater demand as higher temperatures increase evapotranspiration and crop water requirements (Hargreaves and Samani, 1985; Hrozencik et al., 2021). Higher rates of growing season precipitation decrease water demand since precipitation and groundwater applied as irrigation are roughly substitutes.

Table A.1: **Average Treatment Effect IV Model, Full Results**

	Dependent variable:		
	Log(Annual Groundwater Use)		
	(1)	(2)	(3)
Treatment	-0.0414***	-0.0604***	-0.0576***
	(-0.0605,-0.0223)	(-0.0865,-0.0344)	(-0.0858,-0.0294)
$Hi-Use^{own/op}$	-0.0033	-0.0204*	0.0040
	(-0.0179,0.0113)	(-0.0367,-0.0042)	(-0.0153,0.0232)
Treatment X $Hi-Use^{own/op}$		0.0509**	
		(0.0166,0.0852)	
$Hi-Use^{well}$			-0.0197
			(-0.0394,0.00001)
Treatment X $Hi-Use^{well}$			0.0400*
			(0.0056,0.0744)
Temperature	0.0686***	0.0685***	0.0685***
	(0.0653,0.0719)	(0.0651,0.0718)	(0.0652,0.0718)
Precipitation	-0.0227***	-0.0229***	-0.0229***
	(-0.0242,-0.0213)	(-0.0243,-0.0214)	(-0.0243,-0.0214)
Blocking Variables	Yes	Yes	Yes
Well Random Effect	Yes	Yes	Yes
Observations	66,292	66,292	66,292
R^2	0.4509	0.4510	0.4527
F Statistic	684,255.8000***	684,567.6000***	693,794.0000***
Note:	* p<0.05; ** p<0.01; *** p<0.001 Standard errors clustered at the well-owner/operator level. 95 percent confidence intervals in parentheses below coefficient estimates.		

Table A.2: **Average Treatment Effect IV Within-irrigator Spillover Model, Full Results**

	Dependent variable:		
	Log(Annual Groundwater Use)		
	(1)	(2)	(3)
Treatment	-0.0445**	-0.0575**	-0.0623**
	(-0.0745,-0.0144)	(-0.1011,-0.0139)	(-0.1018,-0.0227)
$Hi-Use^{own/op}$	0.0040	-0.0058	0.0061
	(-0.0184,0.0265)	(-0.0295,0.0178)	(-0.0215,0.0338)
Treatment X $Hi-Use^{own/op}$		0.0318	
		(-0.0249,0.0886)	
$Hi-Use^{well}$			-0.0111
			(-0.0484,0.0263)
Treatment X $Hi-Use^{well}$			0.0422
			(-0.0128,0.0973)
Temperature	0.0871***	0.0870***	0.0871***
	(0.0790,0.0953)	(0.0789,0.0952)	(0.0788,0.0954)
Precipitation	-0.0162***	-0.0163***	-0.0163***
	(-0.0195,-0.0130)	(-0.0196,-0.0131)	(-0.0196,-0.0130)
Blocking Variables	Yes	Yes	Yes
Well Random Effect	Yes	Yes	Yes
Observations	39,783	39,783	39,783
R^2	0.3767	0.3768	0.3773
F Statistic	385,887.1000***	386,176.3000***	387,894.8000***
Note:	* p<0.05; ** p<0.01; *** p<0.001 Standard errors clustered at the well-owner/operator level. 95 percent confidence intervals in parentheses below coefficient estimates.		

B Intervention Effects on Annual Groundwater Use and Depletion

The well documented depletion of the High Plains aquifer constitutes a significant threat to the resilience of the agricultural communities that rely on the aquifer's resources and the broader U.S. agricultural economy (Scanlon et al., 2012; Steward et al., 2013; Haacker et al., 2015; Suárez et al., 2018). To understand how our social comparison field experiment affects broader issues related to groundwater sustainability, we estimate the aggregate impact of the intervention on groundwater use and compare this to estimated High Plains aquifer depletion rates in the literature. This analysis demonstrates the extent to which behavioral nudges can be leveraged to address groundwater depletion issues.

We calculate the reduction in aggregate groundwater use induced by the intervention by applying the estimated average treatment effect (see column (1) of table 2 of the main text) to control group well-level average water use during experiment time period. We use the control group as the benchmark for calculating the water use impacts of the intervention as these wells provide the best estimate of counterfactual water use absent the comparison intervention. We aggregate these estimated counterfactual changes in in well-level water use in response to the intervention across all control group wells which yields a predicted aggregate annual reduction of 27,229 acre-feet. Colorado and Kansas treatment wells reduced water use by 8,720 and 18,509 acre-feet per year, respectively (8,720 + 18,509 = 27,229). To calculate the well retirement equivalent of the water conservation induced by the intervention we find the median of the previously calculated well-level average water use during the pre-experiment time period (2011-2018). The median well used 167.4 acre-feet per year, retiring approximately 160 of these median wells results yields an equivalent level of annual groundwater use reduction as that achieved by the intervention in a given year. Note that these predicted aggregate changes in water use represent the impact of the social comparison intervention, they do not correspond to the scenario where all Kansas and Colorado wells receive comparison information.

The previous estimates of water conservation only take into account wells that received comparison information. However, as detailed in section 5.1 of the main text, the intervention also altered pumping behavior among wells owned/operated by treatment group individuals/entities that did not receive comparison information. If we include the change in groundwater use estimated for these wells, then the aggregate reduction in annual groundwater use increases to 37,943 acre-feet which is equivalent to the retirement of approximately 225 median wells. Including these control group wells which would not receive comparison information even if their owner/operator were assigned to the treatment group increases state-level reductions in annual groundwater use in Colorado and Kansas to 10,929 and 27,014 acre-feet, respectively.

We leverage estimates from the literature regarding annual rates of depletion for the High Plains aquifer and compare these volumes to the reductions in groundwater use induced by the intervention. Specifically, we use state-level estimates from Steward and Allen (2016) of annual groundwater depletion rates as of 2020 (see figure 7 in Steward and Allen (2016)) and compare to estimates of intervention induced state level reductions in annual groundwater use. Steward and Allen (2016) estimate Colorado's and Kansas's annual rate of groundwater depletion as of 2020 to be 526,964 and 1,783,570 acre-feet, respectively. Comparing these annual rates of depletion to estimated state-level reductions in groundwater use induced by the intervention reveals that the social comparison reduced annual depletion rates by 1.65% in Colorado and 1.03% in Kansas (8,720/526,964 = 0.0165 & 18,509/1,783,570= 0.0103). When including the interventions' spillover effects on treatment group wells that did not receive comparison information, the reductions in annual depletion increase to 2.07% and 1.51% in Colorado and Kansas, respectively (10,929/526,964 = 0.0207 & 27,014/1,783,570 = 0.0151).

Finally, policymakers may be more interested in the water use impacts of comparison interventions targeting all resource users rather than a subset of users. To that end, we calculate the change in aggregate water use for the scenario where all Kansas and Colorado wells receive comparison information. To do this, we leverage average well-level pre-experiment water use for both treatment and control group wells to establish counterfactual water use in the absence of treatment and apply the estimated treatment effect to predict well-level changes in water use in response to the intervention. We then aggregate these well-level changes in water use across all wells (both treatment and control groups) to generate total changes in water use. This process predicts that if all Kansas and Colorado well owners received comparison information for all their wells, then total water use would decrease by 19,981 and 58,792 acre-feet per year in Colorado and Kansas, respectively (total = 78,774 acre-feet). These predicted aggregate changes in groundwater use would result in a reduction of annual groundwater depletion of 3.7% and 3.2% in Colorado and Kansas, respectively (19,981/526,964 = 0.0379 & 58,792/1,783,570 = 0.0329).

C Empirical Modeling Robustness Checks

This section further tests the robustness of the experimental results presented in section 5 of the main text. Specifically, this section does the following: 1) estimates a model which treats the well-level effects as fixed effects rather than random effects, 2) separately estimates average treatment effects for the Colorado and Kansas experiments to understand how pooling data from both experiments in our analysis of treatment effects impacts results, 3) estimates two models which test the impact of over-treatment, over-control, and treatment-control group contamination on our primary results, 4) estimates treatment effects using a cross-sectional rather than panel model approach, 5) estimates an intent to treat model to test the robustness of our results to the IV approach we leverage to address treatment non-compliance, and 6) estimates a model to assess heterogeneity in treatment effects across time.

C.1 Well-Level Fixed Effects

The primary model used to evaluate the impact of the social comparison intervention leverages a well-level random effect specification. Here we test the robustness of our results to this modeling approach by estimating a similar econometric model to that presented in equation 1 of the main text with a well-level fixed effect. Table C.1 presents the results of this fixed effect specification and exhibits qualitatively similar results to those generated using a well-level random effect. Specifically, the fixed effect specification indicates a negative and statistically significant average treatment effect with a slightly smaller magnitude (approximately 3%) than that estimated with the random effect specification. We also find qualitatively similar results regarding the moderating impact of pre-experiment water use on treatment effects. Overall, results of the fixed effect specification demonstrates that the estimated efficacy of the intervention in reducing subsequent water use persists under differing modeling assumptions related to well-level effects.

Table C.1: **Average Treatment Effect IV Model, Fixed Effect Specification**

	Dependent variable:		
	Log(Annual Groundwater Use)		
	(1)	(2)	(3)
Treatment	-0.0295**	-0.0440**	-0.0360*
	(-0.0490,-0.0101)	(-0.0705,-0.0174)	(-0.0646,-0.0074)
$Hi-Use^{own/op}$	-0.0218**	-0.0348***	-0.0130
	(-0.0366,-0.0070)	(-0.0512,-0.0184)	(-0.0324,0.0064)
Treatment X $Hi-Use^{own/op}$		0.0387*	
		(0.0039,0.0735)	
$Hi-Use^{well}$			-0.0193
			(-0.0391,0.0005)
Treatment X $Hi-Use^{well}$			0.0216
			(-0.0130,0.0562)
Temperature	0.0693***	0.0692***	0.0692***
	(0.0659,0.0726)	(0.0658,0.0725)	(0.0658,0.0725)
Precipitation	-0.0228***	-0.0229***	-0.0229***
	(-0.0242,-0.0213)	(-0.0243,-0.0215)	(-0.0243,-0.0214)
Blocking Variables	Yes	Yes	Yes
Well Fixed Effect	Yes	Yes	Yes
Observations	66,292	66,292	66,292
R^2	0.1202	0.1202	0.1202
F Statistic	7,881.7040***	7,889.6650***	7,886.2610***
Note:	* p<0.05; ** p<0.01; *** p<0.001 Standard errors clustered at the well-owner/operator level. 95 percent confidence intervals in parentheses below coefficient estimates.		

C.2 Separately Modeling Impacts of Colorado and Kansas Interventions

The primary analysis of the comparison intervention's effect on groundwater use (see table 2 in the main text) jointly estimates treatment effects for both the Colorado and Kansas experiments in one unified modeling framework. However, the Colorado and Kansas experiments were conducted independently of each other and differed somewhat in their design and implementation (see section 3 for more information). To test the robustness of our results to the modeling approach utilized in the main text, we separately estimate average treatments effects for both the Colorado and Kansas experiments. Tables C.2 and C.3 present these results for Colorado and Kansas, respectively.

Results of the Colorado experiment qualitatively align with those presented in the main text but are relatively less precisely estimated (see table C.2) . Specifically, results of the Colorado experiment indicate that treatment induces a reduction in subsequent groundwater pumping but this relationship is only statistically significant in the conditional treatment effect specification which includes an interaction between treatment and the well-level pre-experiment water use indicator variable. The relative lack of precision of the treatment effect estimates generated using the Colorado experiment may be related to the diminished statistical power associated with the smaller sample size.

Results of the Kansas experiment also qualitatively align with the pooled results presented in the main text and have a similar level of precision. The Kansas-specific average treatment effect presented in table C.3 is slightly larger in magnitude than that presented in the main text suggesting that treatment effects may be larger in Kansas, potentially due to differences in the experiment's design in Kansas (e.g., providing comparison information in acre inches per acre irrigated rather than total acre-feet). Similar to the results presented in the main text, estimated conditional treatment effects for the Kansas experiment show that treatment effects attenuate but remain negative based on pre-experiment water use patterns.

Table C.2: **Average Treatment Effect IV Model, Colorado Intervention**

	Dependent variable:		
	Log(Annual Groundwater Use)		
	(1)	(2)	(3)
Treatment	-0.0371	-0.0504	-0.0808**
	(-0.0743,0.0001)	(-0.1041,0.0033)	(-0.1361,-0.0255)
$Hi-Use^{own/op}$	0.0169	0.0061	-0.0119
	(-0.0079,0.0418)	(-0.0195,0.0316)	(-0.0445,0.0206)
Treatment X $Hi-Use^{own/op}$		0.0333	
		(-0.0296,0.0961)	
$Hi-Use^{well}$			0.0320
			(-0.0027,0.0667)
Treatment X $Hi-Use^{well}$			0.0657*
			(0.0054,0.1259)
Temperature	0.0721***	0.0722***	0.0721***
	(0.0655,0.0788)	(0.0656,0.0788)	(0.0654,0.0788)
Precipitation	-0.0225***	-0.0225***	-0.0225***
	(-0.0251,-0.0199)	(-0.0251,-0.0199)	(-0.0251,-0.0198)
Blocking Variables	Yes	Yes	Yes
Well Random Effect	Yes	Yes	Yes
Observations	17,552	17,552	17,552
R^2	0.3937	0.3936	0.4038
F Statistic	337,443.9000***	337,178.7000***	402,155.0000***
Note:	* p<0.05; ** p<0.01; *** p<0.001 Standard errors clustered at the well-owner/operator level. 95 percent confidence intervals in parentheses below coefficient estimates.		

Table C.3: **Average Treatment Effect IV Model, Kansas Intervention**

	Dependent variable:		
	Log(Annual Groundwater Use)		
	(1)	(2)	(3)
Treatment	-0.0478^{***}	-0.0719^{***}	-0.0602^{***}
	(-0.0697,-0.0259)	(-0.1007,-0.0430)	(-0.0922,-0.0283)
$Hi-Use^{own/op}$	-0.0060	-0.0287^{**}	0.0044
	(-0.0242,0.0122)	(-0.0496,-0.0079)	(-0.0183,0.0271)
Treatment X $Hi-Use^{own/op}$		0.0661^{**}	
		(0.0258,0.1065)	
$Hi-Use^{well}$			-0.0271^{*}
			(-0.0502,-0.0039)
Treatment X $Hi-Use^{well}$			0.0380
			(-0.0038,0.0799)
Temperature	0.0675^{***}	0.0673^{***}	0.0673^{***}
	(0.0638,0.0713)	(0.0636,0.0711)	(0.0636,0.0711)
Precipitation	-0.0228^{***}	-0.0231^{***}	-0.0230^{***}
	(-0.0245,-0.0212)	(-0.0247,-0.0214)	(-0.0247,-0.0214)
Blocking Variables	Yes	Yes	Yes
Well Random Effect	Yes	Yes	Yes
Observations	48,740	48,740	48,740
R^2	0.4623	0.4625	0.4627
F Statistic	$439{,}845.8000^{***}$	$440{,}317.5000^{***}$	$441{,}013.7000^{***}$
Note:	* p<0.05; ** p<0.01; *** p<0.001 Standard errors clustered at the well-owner/operator level. 95 percent confidence intervals in parentheses below coefficient estimates.		

C.3 Duplicated Addresses and Multiple Mailers

Section 3 of the main text describes several of the experimental design challenges encountered during the implementation of the social comparison intervention. Here we test the robustness of our results to the challenges associated with variations in abbreviation rules in the address data used to identify experiment units (i.e., well owners/operators with a common address). These address issues resulted in some well owners/operators showing up multiple times in either the treatment or control groups. To assess the extent to which over-treatment (an individual well owner/operator receiving multiple mailers) or over-control (an individual well owner/operator being in the control group multiple times), we estimate average treatment effects using a sample of data which drops these entities. Results of this model are presented in table C.4 whose estimated average treatment effects are similar in magnitude as those presented in the main text. These results suggest that our estimated treatment effects remain robust to over-control and over-treatment issues.

In some cases, address issues resulted in well owners/operators being classified in both the treatment and control group. The model estimating average treatment effects in the main text drops well observations associated with these treatment-control group well owners/operators. Here we test the extent to which dropping these observations affects the estimated treatment effects of the intervention. To do so, we estimate a model which includes these treatment/control group observations and address discrepancies between control group designation and treatment receipt using IV non-compliance methods outlined in Angrist et al. (1996). While the treatment-control issue does not perfectly align with Angrist et al.'s (1996) non-compliance framework, results of this model provide some intuition regarding the impact of treatment-control group duplication on our results. Table C.5 presents results of this model which are similar in magnitude to those presented in the main text providing evidence that treatment-control group issues do not significantly impact our primary results.

Table C.4: **Average Treatment Effect IV Model, Dropping Over-control and Over-treatment Observations**

	Dependent variable:		
	Log(Annual Groundwater Use)		
	(1)	(2)	(3)
Treatment	-0.0398***	-0.0603***	-0.0588***
	(-0.0598,-0.0198)	(-0.0876,-0.0330)	(-0.0882,-0.0294)
$Hi-Use^{own/op}$	-0.0030	-0.0219*	0.0049
	(-0.0185,0.0124)	(-0.0390,-0.0047)	(-0.0153,0.0250)
Treatment X $Hi-Use^{own/op}$		0.0554**	
		(0.0193,0.0914)	
$Hi-Use^{well}$			-0.0223*
			(-0.0429,-0.0017)
Treatment X $Hi-Use^{well}$			0.0468*
			(0.0108,0.0828)
Temperature	0.0681***	0.0680***	0.0680***
	(0.0647,0.0716)	(0.0645,0.0714)	(0.0646,0.0714)
Precipitation	-0.0230***	-0.0231***	-0.0231***
	(-0.0245,-0.0215)	(-0.0246,-0.0216)	(-0.0246,-0.0216)
Blocking Variables	Yes	Yes	Yes
Well Random Effect	Yes	Yes	Yes
Observations	62,352	62,352	62,352
R^2	0.4419	0.4420	0.4438
F Statistic	625,504.4000***	625,726.3000***	634,180.6000***
Note:	* p<0.05; ** p<0.01; *** p<0.001 Standard errors clustered at the well-owner/operator level. 95 percent confidence intervals in parentheses below coefficient estimates.		

Table C.5: **Average Treatment Effect IV Model, Including Treatment-control Observations**

	Dependent variable:		
	Log(Annual Groundwater Use)		
	(1)	(2)	(3)
Treatment	-0.0466***	-0.0692***	-0.0631***
	(-0.0692,-0.0241)	(-0.1008,-0.0375)	(-0.0974,-0.0288)
$Hi-Use^{own/op}$	-0.0025	-0.0220**	0.0047
	(-0.0169,0.0118)	(-0.0373,-0.0068)	(-0.0139,0.0234)
Treatment X $Hi-Use^{own/op}$		0.0557**	
		(0.0210,0.0903)	
$Hi-Use^{well}$			-0.0224*
			(-0.0413,-0.0035)
Treatment X $Hi-Use^{well}$			0.0405*
			(0.0054,0.0755)
Temperature	0.0687***	0.0685***	0.0686***
	(0.0655,0.0718)	(0.0653,0.0717)	(0.0654,0.0718)
Precipitation	-0.0226***	-0.0228***	-0.0228***
	(-0.0240,-0.0212)	(-0.0242,-0.0214)	(-0.0242,-0.0214)
Blocking Variables	Yes	Yes	Yes
Well Random Effect	Yes	Yes	Yes
Observations	73,158	73,158	73,158
R^2	0.4603	0.4542	0.4516
F Statistic	755,582.2000***	723,752.0000***	710,449.3000***
Note:	* p<0.05; ** p<0.01; *** p<0.001 Standard errors clustered at the well-owner/operator level. 95 percent confidence intervals in parentheses below coefficient estimates.		

C.4 Cross-Sectional Econometric Modeling Results

The primary econometric model used to analyze the impact of the comparison treatment on subsequent water use leverages a panel data approach to estimate average treatment effects. This section of the appendix presents empirical results generated using a cross-sectional approach to estimate average treatment effects. These cross-section models use an econometric model similar to that described by equation 1 in the main text that excludes the well-level random effect and separately estimates state-year models i.e., separate models are estimated for Colorado's 2019, 2020, and 2021 experiments and Kansas's 2020 and 2021 experiments. These results aim to check the robustness of our results to differing modeling specifications and align with early iterations of the experiment's pre-analysis plan which outlined cross-sectional models to estimate the impact of the intervention.

Tables C.6 and C.7 present these cross-sectional modeling results for the Colorado and Kansas experiments, respectively. Similar to the panel model results in the main text, results presented in tables C.6 and C.7 use an instrumental variables approach to address treatment non-compliance (Angrist et al., 1996). Each column of tables C.6 and C.7 present results from differing experiment years within a given state. For simplicity, we do not estimate conditional treatment effects using the cross-section approach. Colorado and Kansas cross-sectional results find similar magnitude treatment effects to those presented in the main text using a panel data approach. However, treatment effects are only statistically significant for the Kansas experiment. The lack of statistically significant treatment effects for the Colorado experiment may be related to a relative lack of statistical power to detect effects when relying on the relatively smaller sample of wells involved in a single iteration of the Colorado experiment. Estimated coefficients related to other model covariates also generally align with panel data results (see tables A.1 and A.2). An exception is the estimated impact of temperature on groundwater use in Kansas where cross-section results suggest a negative, although statistically insignificant, relationship. The lack of precision for these estimated temperature coefficients may be related to relatively little within year variation between wells in average growing season temperature in Kansas.

Table C.6: **Average Treatment Effect IV Cross-sectional Model, Colorado**

	Dependent variable:		
	Log(Annual Groundwater Use)		
	(2019)	(2020)	(2021)
Treatment	-0.0372	-0.0249	-0.0184
	(-0.1097,0.0352)	(-0.0997,0.0498)	(-0.0892,0.0525)
$Hi-Use^{own/op}$	0.4204***	0.3830***	0.4120***
	(0.3550,0.4858)	(0.3194,0.4467)	(0.3498,0.4743)
Temperature	0.1181*	0.0242	0.0810
	(0.0024,0.2337)	(-0.0987,0.1472)	(-0.0003,0.1622)
Precipitation	-0.0396**	-0.0042	-0.0071
	(-0.0617,-0.0176)	(-0.0365,0.0281)	(-0.0262,0.0120)
Comparison group indicator variables	Yes	Yes	Yes
Observations	1,614	1,614	1,614
R^2	0.2369	0.1812	0.2362

Note: * p<0.05; ** p<0.01; *** p<0.001
Standard errors clustered at the well-owner/operator level.
95 percent confidence intervals in parentheses below coefficient estimates.

Table C.7: **Average Treatment Effect IV Cross-sectional Model, Kansas**

	Dependent variable:	
	Log(Annual Groundwater Use)	
	(2019)	(2020)
Treatment	-0.0598*	-0.0621*
	(-0.1068,-0.0128)	(-0.1135,-0.0107)
$Hi-Use^{own/op}$	0.2624***	0.2675***
	(0.2153,0.3096)	(0.2176,0.3173)
Temperature	-0.2697***	-0.1734***
	(-0.3519,-0.1875)	(-0.2556,-0.0912)
Precipitation	-0.0493***	-0.0232*
	(-0.0693,-0.0293)	(-0.0433,-0.0030)
Comparison group indicator variables	Yes	Yes
Observations	4,605	4,605
R^2	0.3470	0.3292
Note:	* p<0.05; ** p<0.01; *** p<0.001 Standard errors clustered at the well-owner/operator level. 95 percent confidence intervals in parentheses below coefficient estimates.	

C.5 Intent to Treat Models

The econometric modeling presented in the main text of the paper addresses treatment non-compliance using an instrumental variable approach first proposed by Angrist et al. (1996). Here we assess the robustness of our results to this modeling approach by estimating an 'intent to treat' model which measures the average causal effect of assignment to the treatment group rather than treatment receipt (Gupta, 2011). Table C.8 presents the results for the 'intent to treat' model demonstrating that estimated average treatment effects do not vary significantly based on this modeling approach. Specifically, the 'intent to treat' model finds an average treatment effect of nearly 4% which aligns with the treatment effect estimated with the IV approach. Additionally, we find similar conditional average treatment effects with the 'intent to treat' model i.e. the impact of treatment is largest among wells and and well owners/operators with lower than average pre-experiment water use while treatment effects attenuate for wells and well owners/operators with higher than average pre-experiment water use.

Table C.8: **Average Treatment Effect Intent to Treat Model**

	Dependent variable:		
	Log(Annual Groundwater Use)		
	(1)	(2)	(3)
Treatment	-0.0383***	-0.0558***	-0.0532***
	(-0.0551,-0.0215)	(-0.0783,-0.0334)	(-0.0763,-0.0301)
$Hi-Use^{own/op}$	-0.0033	-0.0204*	0.0039
	(-0.0177,0.0111)	(-0.0366,-0.0041)	(-0.0152,0.0230)
Treatment X $Hi-Use^{own/op}$		0.0507**	
		(0.0178,0.0836)	
$Hi-Use^{well}$			-0.0196*
			(-0.0392,-0.000002)
Treatment X $Hi-Use^{well}$			0.0399*
			(0.0086,0.0712)
Temperature	0.0686***	0.0685***	0.0685***
	(0.0653,0.0719)	(0.0652,0.0718)	(0.0652,0.0718)
Precipitation	-0.0227***	-0.0229***	-0.0228***
	(-0.0241,-0.0213)	(-0.0243,-0.0214)	(-0.0243,-0.0214)
Blocking Variables	Yes	Yes	Yes
Well Random Effect	Yes	Yes	Yes
Observations	66,292	66,292	66,292
R^2	0.4510	0.4513	0.4532
F Statistic	684,815.4000***	685,765.0000***	695,472.1000***
Note:	* p<0.05; ** p<0.01; *** p<0.001 Standard errors clustered at the well-owner/operator level. 95 percent confidence intervals in parentheses below coefficient estimates.		

C.6 Persistence of Treatment Effects Across Time

The long-term conservation benefits of behavioral nudges like that presented in the main text depend crucially on the persistence of treatment effects over time. If individuals habituate to information provided by nudges over time, then the estimated benefits of the intervention will similarly attenuate over time. In this section of the appendix, we explore the persistence of treatment effects over time in the context of the social comparison intervention presented in the main text. To empirically model temporally heterogeneous treatment effects we introduce two dummy variables, ‘First Treatment’ and ‘Last Treatment,’ which indicate the first or last year of treatment for a given well owned/operated by a treatment group individual/entity. We include these dummy variables and terms interacting them with the treatment variable in two separate models similar to that outlined by equation 1 of the main text. Results of these models are presented in columns (2) and (3) of table C.9. For ease of comparison, column (1) of table C.9 displays the primary average treatment effect results from the main text.

Table C.9: **Average Treatment Effect IV Model, Temporally Heterogeneous Effects**

	Dependent variable:		
	Log(Annual Groundwater Use)		
	(1)	(2)	(3)
Treatment	-0.0414***	-0.0599***	-0.0297**
	(-0.0605,-0.0223)	(-0.0826,-0.0372)	(-0.0519,-0.0075)
$Hi-Use^{own/op}$	-0.0033	-0.0041	0.0160*
	(-0.0179,0.0113)	(-0.0198,0.0117)	(0.0003,0.0318)
First Treatment		0.0014	
		(-0.0151,0.0180)	
Treatment X First Treatment		0.0366**	
		(0.0109,0.0622)	
Last Treatment			-0.0720***
			(-0.0903,-0.0537)
Treatment X Last Treatment			0.0180
			(-0.0089,0.0449)
Temperature	0.0686***	0.0683***	0.0660***
	(0.0653,0.0719)	(0.0650,0.0717)	(0.0627,0.0694)
Precipitation	-0.0227***	-0.0231***	-0.0246***
	(-0.0242,-0.0213)	(-0.0245,-0.0216)	(-0.0261,-0.0231)
Blocking Variables	Yes	Yes	Yes
Well Random Effect	Yes	Yes	Yes
Observations	66,292	66,292	66,292
R^2	0.4509	0.4513	0.4521
F Statistic	684,255.8000***	686,213.1000***	687,606.2000***
Note:	* p<0.05; ** p<0.01; *** p<0.001 Standard errors clustered at the well-owner/operator level. 95 percent confidence intervals in parentheses below coefficient estimates.		

Column (2) of table C.9 presents empirical results evaluating if treatment effects in the first year of the interventions differs from subsequent years. The estimated coefficient for the ‘Treatment’ variable in column (2) suggests an approximate reduction in groundwater use of 6% after the initial year (‘First Treatment’ = 0). Jointly, the coefficient estimates for ‘Treatment’ and the term interacting ‘Treatment’ with ‘First Treatment’ show that the initial intervention resulted in approximately a 2.5% reduction in groundwater use while subsequent treatment effects increase in later years of the experiment. The conditional treatment effects presented in column (2) provides some evidence suggesting that treatment effects do not attenuate over time but instead is amplified. We further test this hypothesis in column (3) of table C.9 which tests for temporally heterogeneous treatment effects in the last year of the intervention. The estimated coefficient for the 'Treatment' variable in column (3) suggest that the first year of the intervention induced approximately a 3% reduction in groundwater use. The coefficient estimate for the term interacting 'Treatment' with the 'Last Treatment' indicator is positive but not statistically significant which suggests that treatment effects are not significantly different in the last year of the intervention. This result provides some evidence that treatment group individuals/entities do not habituate to comparison information over time, at least within the relatively short time span (2 or 3 years) evaluated in this experiment.

D Intervention Costs Per Unit of Water Conserved

In this section of the appendix we develop a simple methodology to calculate the per unit of water conserved costs of the social comparison intervention. This calculation facilitates cost-effectiveness comparisons between the intervention and other water conservation policies implemented in the study areas, namely well retirement programs. Determining the per unit of water conserved cost of the intervention requires accounting for both the resources necessary to generate and mail the intervention and the change in water use behavior attributable to the intervention.

Colorado and Kansas mailers cost approximately $1.80 and $1.20 per mailer, respectively. The Kansas mailers were somewhat less expensive per mailer due to quantity discounts applied to larger mailing and printing jobs. The total cost per year for the Colorado and Kansas interventions were $876 and $1,667, respectively. Table D.1 outlines the total costs and per mailer costs, differentiating between printing and mailing for the Colorado and Kansas interventions. Note that the number of mailers sent in both experiments is slightly larger than the treatment group as mailers were also sent to the research team and some treatment mailers were excluded due to address non-compliance in the first years of the interventions.

Table D.1: Social Comparison Intervention Mailing and Printing Costs Per Year

		Total Cost ($)	Number of Mailers	Cost per Mailer ($)
Colorado	Printing	556	494	1.12
	Mailing	320	494	0.65
	Total	876	494	1.77
Kansas	Printing	901	1418	0.64
	Mailing	766	1418	0.54
	Total	1667	1418	1.18

To calculate intervention costs per acre-foot conserved we calculate aggregate annual groundwater use reductions arising from the intervention and compare this to total annual intervention costs. We calculate the reduction in aggregate groundwater use induced by the intervention by applying the estimated average treatment effect (see column (1) of table 2 of the main text) to the control group well-level average water use during the experiment time period. We use the control group as the benchmark for calculating the water use impacts of the intervention as these wells provide the best estimate of counterfactual water use absent the comparison intervention. We aggregate predicted changes in well-level water use in response to the intervention across all control group wells which yields an aggregate annual reduction of 27,229 acre-feet, with Colorado and Kansas wells reducing water use by 8,720 and 18,509 acre-feet per year, respectively (8,720 + 18,509 = 27,229). Using the total annual intervention costs reported in table D.1 implies average annual per acre-foot conserved cost of $0.10 and $0.09 in Colorado and Kansas, respectively ($876/8,720 acre-feet = 0.1004 $ per acre-foot & $1,667/18,509 acre-feet = 0.0901 $ per acre-foot).