Reducing carbon emissions of households through monetary incentives and behavioral interventions: a meta-analysis

Tarun Khanna[1,5], Giovanni Baiocchi[3], Max Callaghan[1,2], Felix Creutzig[1,4], Horia Guias[6], Neal R Haddaway[1,7,8], Lion Hirth[5,9], Aneeque Javaid[1], Nicolas Koch[1], Sonja Laukemper[1], Andreas Löschel[6], Maria Del Mar Zamora[1], Jan C Minx[1,2*]

Despite the importance of evaluating all mitigation options so as to inform policy decisions addressing climate change, a systematic analysis of household-scale interventions to reduce carbon emissions is missing. Here, we address this gap through a state-of-the-art machine-learning assisted meta-analysis to comparatively assess the effectiveness of a range of monetary and behavioral interventions in energy demand of residential buildings. We identify 122 studies and extract 360 effect sizes representing trials on 1.2 million households in 25 countries. We find that all the studied interventions reduce energy consumption of households. Our meta-regression evidences that monetary incentives are on an average more effective than behavioral interventions, but deploying the right combinations of interventions together can increase overall effectiveness. We estimate global cumulative carbon emissions reduction of 8.64 Gt CO_2 *by 2040, though deploying the most effective packages and interventions could result in greater reduction. While modest, this potential should be viewed in conjunction with the need for de-risking mitigation with energy demand reductions and realizing substantial co-benefits.*

[1] Mercator Research Institute on Global Commons and Climate Change, Torgauer Straße 12–15, EUREF Campus #19, 10829 Berlin, Germany
[2] School of Earth and Environment, University of Leeds, Leeds LS2 9JT, United Kingdom
[3] Department of Geographical Sciences, University of Maryland, USA
[4] Technische Universität Berlin, Berlin, Germany
[5] Hertie School, Berlin, Germany
[6] Center for Applied Economic Research, University of Münster, Münster, Germany
[7] Stockholm Environment Institute, Linnégatan 87D, Stockholm, Sweden
[8] Africa Centre for Evidence, University of Johannesburg, Johannesburg, South Africa
[9] Neon Neue Energieökonomik

*Corresponding author, minx@mcc-berlin.net

Finding low energy demand pathways is necessary to hedge against the risks involved in decarbonizing energy supply and is key for finding socially acceptable ways of meeting the Paris climate goals[1–4]. Energy demand from buildings was responsible for 28% of global energy-related CO_2 emissions in 2019, when indirect emissions from upstream power generation are considered. In absolute terms, buildings-related CO_2 emissions increased to an all-time high of 10 $GtCO_2$, with residential buildings accounting for 60% of these emissions[5]. According to the IEA, this new trend contrasts with the plateauing of emissions from 2013 to 2016. Since then increased demand for building energy services has outpaced energy efficiency and de-carbonization efforts[6]. Besides technologies and architecture, behavior, lifestyle, and culture have a major effect on buildings' energy demand with three to five-fold difference in energy use for provision of similar building-related energy service levels[7]. A lack of systematic efforts to quantify demand side solutions, in general, and interventions in household energy demand specifically, has led to a bias towards riskier supply side solutions in climate change assessments such as those by the Intergovernmental Panel on Climate Change[8].

There is a rich and diverse literature available on demand-side solutions[9]. Since the oil price shock in 1970s, interventions to reduce energy use in building and appliance use have been researched extensively[10]. Experiments that use monetary incentives to reduce consumption have been trialed widely, even more so since the introduction of smart metering at scale over the last decade[11]. Evidence has accumulated on use of behavioral interventions, which encompass a range of initiatives that may, either by themselves or in conjunction with the more typical policy tools (e.g., infrastructure, incentives), achieve greater energy consumption reductions than have been achieved by the typical tools alone[12]. In spite of this vast evidence pool that can be employed for policymaking, little is known about the global carbon emissions reduction potential of such interventions.

We address this gap through an interdisciplinary meta-analysis of interventions in household energy consumption. Previous reviews tend to be disciplinary and focused on subsets of the interventions. Faruqui et al.[13] on pricing interventions, Karlin et al.[14] on feedback, Abrahamse et al.[15] and Andor et al.[16] on social comparison, commitment devices, goal setting, and labelling. Nisa et al.[17] consider evidence from a wider range of household behaviors that are relevant for climate change mitigation but did not review interventions in energy consumption exhaustively. The meta-analysis by Delmas et al.[18] broke new ground but was based on a narrower literature search and does not include studies published after 2012, which constitute about half of our sample. This paper provides a comprehensive, up-to-date meta-analysis that critically assesses energy savings potential of pecuniary and behavioral interventions in household energy consumption, as well as their carbon implications to inform upcoming climate change assessments.

We extend previous analyses in important ways. First, following international standards for systematic reviews[19], we do not restrict our literature search based on research design, source or timelines. The resulting sample of relevant studies is at least twice as large as previous analyses, which allows us to run rigorous multilevel meta-analysis models to increase reliability of results. Second, with the exponential growth of the literature in recent years, there are several new studies from countries like Japan, China, India, Israel, and Australia that provide regionally varied insights. Third, none of the previous reviews estimate mitigation potentials, the commonly applied metric in climate change assessments. We translate the evidence on interventions in energy consumption into meaningful estimates of CO_2 reduction potentials. Last, all the information collected, and code developed in this project is publicly available in line with the systematic reviews reporting protocol (ROSES)[25,26], providing the transparency and reproducibility required to conform with Open Synthesis[22] principles.

Interventions targeting household energy consumption

We perform a systematic review and meta-analysis of the literature (see methods) on interventions in residential energy demand. These interventions can broadly be grouped into monetary incentives that offer households a tangible financial reward for reducing energy consumption, and behavioral interventions that include altering decision environments (often referred to as choice architecture) or nudging, appealing to norms, providing easily interpretable and credible information at the point of decision-making, and improving skills required to perform or forego behaviours[12]. Following previous studies[14,16,18] we classify behavioral interventions into information, feedback, social norms and motivation interventions. We systematically search, screen and select the relevant literature on the five different types of interventions (see Figure 1).

Figure 1 Typology of reviewed interventions

Intervention type	Intervention	Description
Monetary Incentives	Critical Peak/ Seasonal Pricing Time of Use/ Real-time Pricing Rewards/ Rebates	Time of use pricing aligns the prices faced by households with the underlying cost of supply, which is higher during peak demand periods[23]. Other interventions reward consumers for reducing peak period consumption[24]. Households are expected to reduce consumption as long as the financial savings from reduced consumption outweigh the costs of shifting or reducing consumption[25].
Information	Home Audits Tips Reminders	These policies focus on promoting energy saving behavior by reducing the information deficit faced by households with activities and actions that can help reduce energy consumption[15]. The information provided may be general advice like energy saving tips and practices through workshops[26] and mass media campaigns[27] or tailored advice in the form of home audits[28].
Feedback	Historical In-home displays	Feedback interventions are rooted in psychological research that posits that directing an individuals' attention to a feedback-standard gap that is relevant to the individuals can engender behavioral change[14]. Most experiments provide individuals information about their energy use, drawing comparisons to the historical consumption[29]. The effect of feedback seems to depend on its frequency, medium and duration[14,30].
Social Comparison	Home energy reports Normative feedback	Households are benchmarked against the performance of their social group[18,31]. Norm based communication has been widely adopted by utilities in the form of Home Energy Reports[32], which seem to be effective in some cases even years after households received their initial reports[33].
Motivation	Commitment Devices Goal Setting Gamification	Social pressure has also been employed in the form of public pledges or commitments by households to practice energy conserving behaviours[34]. Goal setting interventions in which households commit to reducing energy consumption by a certain percentage over the course of the experiment are other commitment devices[16]. Some recent experiments have used web based gamified platforms or mobile apps to induce behavioral change.

We ultimately identify and code 122 relevant studies across disciplines and geographies. This is twice the number of studies included in previous meta-analyses (see methods, SI). We extract 360 effect sizes from these studies, or an average of about three effects per study. Our final sample represents research on a total of 1.2 million households across 25 countries. About half of the sample comes from studies in economics or business, about a quarter from psychology and around a fifth from engineering or technology literature. The earliest studies date back to the mid-1970s, but around half of the sample is from studies conducted after 2013. About 45% of the sample comes from households in the United States, 25% in continental Europe, and another 10% in the United Kingdom. The number of studies looking at Asian households is increasing recently and constitutes 10% of the sample with the remaining 10% coming from Australia, Latin America, Africa, and the Middle East. The mean (standard deviation) baseline consumption across effects is 7439 (8845) kWh yr^{-1} and the mean duration of the underlying experiments is 21.5 (26.8) weeks.

The studies in our sample reported effects in terms of relative change in energy consumption but the exact dependent variable and statistical technique employed (various regression models, difference of means, etc.) vary across studies. In order to estimate the aggregate effect size, we first standardized the effects by converting the estimates reported by each study to Fisher's Z[35] and then used meta-analysis models to calculate the aggregate effect across studies (see methods).

Table 1 Descriptive statistics of the sample of included studies

	No. of effects	% of total sample	% Reduction in Energy Consumption			Standardized Effect Size (Z)	
			Average	Standard Deviation	Weighted Average	Average	Standard Deviation
Feedback	26	63.1	5.04	6.91	1.77	0.133	0.213
Information	174	48.6	5.61	6.84	1.91	0.166	0.245
Monetary Incentives	75	20.9	6.06	6.41	1.44	0.148	0.188
Motivation	73	20.4	9.51	9.73	1.87	0.187	0.161
Social Comparison	134	37.4	5.34	7.62	1.81	0.131	0.208
All Interventions	360		5.83	7.41	1.80	0.149	0.214

What interventions work best

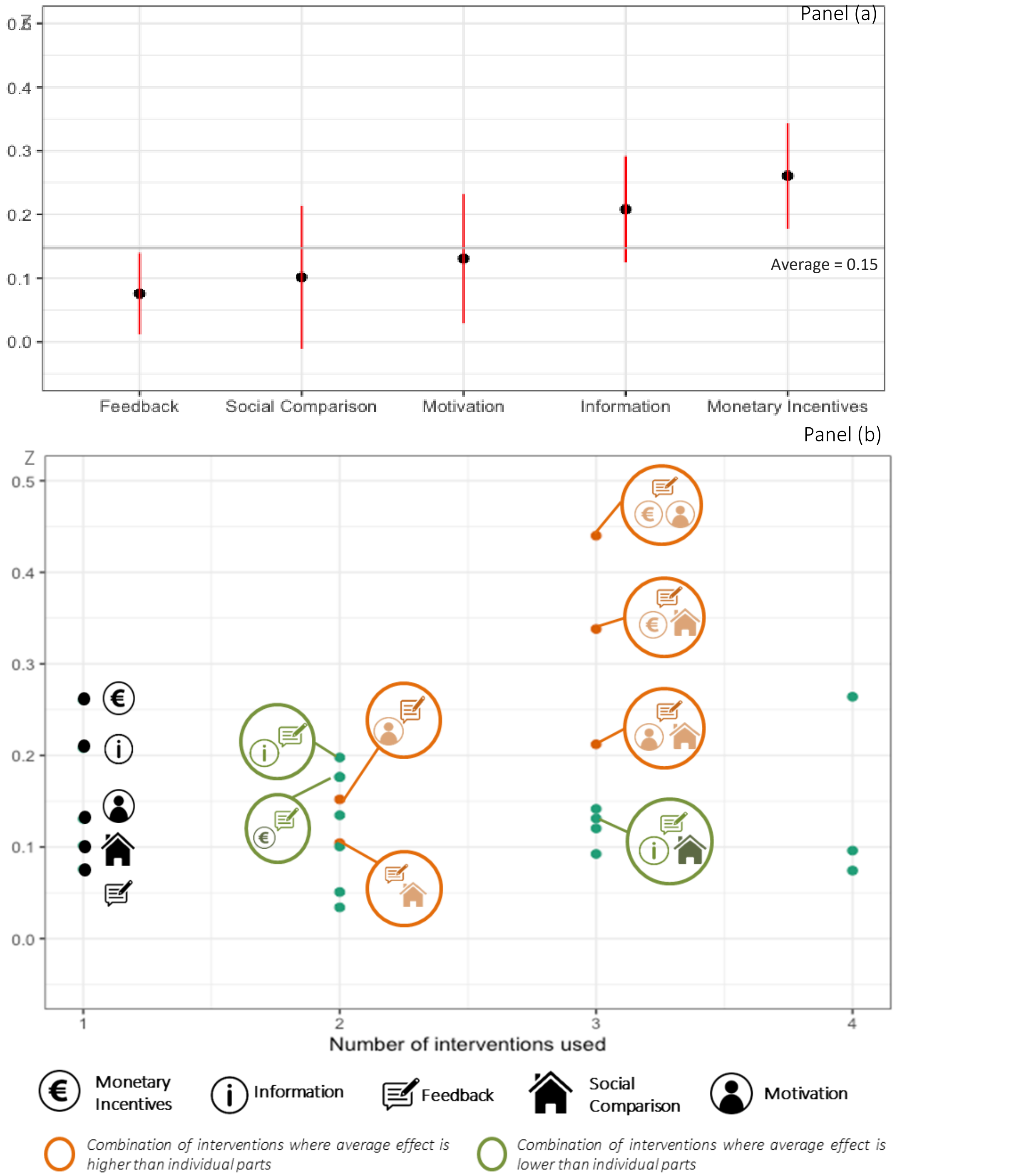
Our analysis finds a medium average effect size across all interventions. The estimated average effect varies between 0.10 – 0.15 and is both statistically significant and substantive across model specifications. The average effect size is 0.10 (95% CI = [0.08, 0.11]; 95% prediction interval = [0.02, 0.18]) in a random effects model with DerSimonian-Laird (DL) estimator and 0.15 (95% CI = [0.13, 0.17]; 95% prediction interval = [-0.23, 0.53]) with a random effects model with Restricted Maximum Likelihood (REML) estimator. The REML estimator is recommended when the heterogeneity is large, as is in our sample[36]. We also estimated a multilevel model to account for dependence between effect sizes coming from the same studies. This gives an average effect size of 0.15 (95% CI = [0.12, 0.18]; 95% prediction interval = [-0.22, 0.52]). These estimates are consistent with the re-examination of data collected by Nisa et al.[16, 35]. The results are robust to influential study analysis and variance matrix specification (see methods). While an average effect size of 0.10 can still be considered small at the level of a single household intervention but relevant if scaled up, an average effect size of 0.15 indicates a medium effect and is considered to be consequential both at a single household level and cumulative over many households[37,38].

Our analysis reveals distinct differences in average effect sizes across individual interventions (Figure 2a). Studies that solely focused on monetary incentives (0.26; 95% CI = [0.17, 0.34]) and information (0.21; 95% CI = [0.13, 0.29]) find higher average effect sizes than studies concerned only with feedback (0.08;

95% CI = [0.01, 0.14]) and social comparison (0.10; 95% CI = [-0.01, 0.21]). The average effect of motivation studies (0.13; 95% CI = [0.03, 0.23]) is close to the overall average though with wider confidence intervals.

We find evidence that combinations of interventions are additive in their effect and may even perform better (Figure 2b). For example, the average effect for studies that combine feedback, social comparison, and monetary interventions is higher (0.33; 95% CI = [0.06, 0.61]) than average effect size for feedback, monetary interventions, social comparison individually. The average effect size of studies, that combine motivation, feedback, and monetary incentives (0.44; 95% CI = [0.09, 0.78]) or motivation, feedback, and social comparison (0.21; 95% CI = [0.01, 0.38]), is also higher than the effect size of individual interventions. In other cases, the overall effect size is about the same as the individual effects; for example in the combination of feedback and social comparison (0.10; 95% CI = [0.01, 0.18]). Interestingly, the average effect from combining feedback and monetary incentives (0.17; 95% CI = [0.06, 0.29]) is lower than the average effect of monetary incentives alone. This supports the trade-off between altruistic and pecuniary motives for reducing energy consumption found in primary studies[25,39,40]. Surprising, there is a similar trend in other combinations involving information, feedback, and social comparison. A Wald-type chi-square test confirms that the differences between the average effect of the combination of interventions noted above and their respective constituents are statistically significant[41]. These results are robust to the choice of model and influential study analysis, though removing influential studies reduces the differences between the various combinations. Overall, while these results support the idea that behavioral interventions should not be looked at only individually but rather as packages to increase effectiveness[12], there might also be trade-offs in certain combinations.

Figure 2 Panel (a) shows the average effect size across interventions along with the 95% confidence intervals. Panel (b) shows average effect size for combination of interventions. Z > 0 implies reduction in energy consumption and Z <= 0 implies increase in energy consumption as a result of the interventions



Notes: The results shown here use the multilevel meta-regression model with interacted dummy variables for the five interventions. In the panel (b) only the combination of interventions for which the average effect size was statistically significant at the 5% level have been labelled.

Explaining heterogeneity in effect sizes

The meta-analysis models used to estimate the aggregate treatment effects indicate a high degree of heterogeneity in effect sizes across studies ($I^2 = 94.12$ for DL model and 99.74 for REML model). In order to understand what drives effect size heterogeneity we performed a meta-regression controlling for a range of study characteristics including region and time of study, study design and a range of study level controls (Table 2).

Household interventions may vary across regions and countries[13,42]. We find that compared to the studies from the United States, average effect in studies done in Asia is higher, especially those that employ monetary incentives. Average effect in studies from continental Europe is marginally larger but the difference is not statistically significant. Overall we do not find significant differences in results reported from different regions.

Our study confirms that the average effect reported by newer studies is lower. We find a statistically significant negative coefficient for the study year moderator in eight of the ten model specifications. We also find that studies with longer treatment duration tend to find smaller effects on average questioning the magnitude and sustainability of induced behavioral changes. The coefficient of treatment duration is negative and statistically significant in five of the model specifications. While the coefficient is not large, it predicts that studies with treatment duration of more than 100 weeks will find negligible effects. However, long term studies are scarce—the mean (median) treatment duration in our sample is only 21.5 (12) weeks, indicating need for long-term trials.

We further find that rigorous study designs find lower effect sizes. The primary studies in our dataset either compared the electricity consumption of the households before and after an intervention (a pre-post design), or across treatment and control groups, or both before and after intervention and across treatment groups (difference in difference design, DID). The control-treatment and DID designs studies on average report lower reduction in energy consumption. The coefficient of the moderator variables are statistically significant and negative when all interventions are considered together, and also for subsets of interventions except motivation and monetary incentives.

Household selection also impacts study outcomes. With monetary incentives, which have largest effect size, the effects are lower for households that did not opt-in into the experiment. The coefficient of moderator variable for opt-in is positive and statistically significant. There are no statistically significant differences in the results between studies that employed randomization and studies that did not, except in case of feedback.

Finally, studies that control for weather have lower average effects, though this difference is not statistically significant except for motivation studies. Studies that control for characteristics of the house (size, appliances) tend to find a smaller effect on average, a finding that is consistent across all model specifications but is statistically significant only for monetary incentives. On the other hand, the moderator variable for demographic differences between the households is inconsistent and not statistically significant.

Table 2 Results from the meta regression model. Dependent variable is Fisher's Z, Z > 0 implies reduction in energy consumption and Z =< 0 implies increase in energy consumption

	All		Feedback		Information		Monetary Incentives		Motivation		Social Comparison	
	REML	Multilevel	REML	Multilevel	REML	Multilevel	REML	Multilevel	REML	Multilevel	REML	Multilevel
Asia	0.14***	0.15**	0.03	0	0.06	0.13	0.22**	0.14	0.05	0.12	0.02	0.01
UK	0.01	0.03	-0.03	-0.01	0.04	0.06	-0.07	-0.13	-0.04	0.09	-0.07	-0.09
Europe excl. UK	0.02	0.05	0.01	-0.01	0.02	0.09	0.16*	0.09	0	0.06	-0.11*	-0.03
Other regions	0.01	0.04	-0.01	0.02	0.02	0.06	-0.06	-0.02	0.10*	0.14	-0.07	-0.08
Study Year	-0.00***	-0.00**	-0.01***	-0.01***	-0.01***	-0.00*	0	0	-0.00*	-0.01*	-0.01*	0
Treatment Period	-0.00**	-0.00*	-0.00***	-0.00*	0	0	-0.00*	0	0	0	0	0
Study Design - DID	-0.17***	-0.11*	-0.28***	-0.12*	-0.25***	-0.20*	0.01	0.05	0.16*	0.17	-0.32***	-0.24**
Study Design Control-treatment	-0.13**	-0.08	-0.29***	-0.11	-0.23***	-0.13	0.16	0.05	0.20**	0.14	-0.33***	-0.25**
Randomization Yes	0.03	0.04	0.10*	0.05	0.03	-0.02	-0.11	-0.06	0.02	-0.09	-0.06	-0.05
Opted In Yes	0.03	0.03	0	0	0.05	0.05	0.22***	0.16	-0.07	-0.05	0.06	0.06
Household Type	-0.02	0.03	0.01	-0.01	0.01	0.03	0.60*	0.46	-0.03	0.03	-0.01	-0.02
Residence Type	-0.03	-0.08	-0.03	-0.01	-0.03	-0.11	-0.64*	-0.41	-0.02	-0.04	-0.05	-0.04
Weather	-0.02	0	-0.02	0	-0.06	-0.02	-0.11	-0.08	-0.09*	0	-0.01	0.02
Intercept	8.72***	8.44**	12.77***	11.71***	12.79***	10.13*	-1.14	0.07	7.53*	13.69*	12.97**	10.35
No. of Effects	324	324	198	198	150	150	71	71	72	72	115	115
I2	99.47		99.23		99.64		96.87		50.42		99.51	
R2	25.52		51.46		29.37		49.04		95.82		47.45	

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

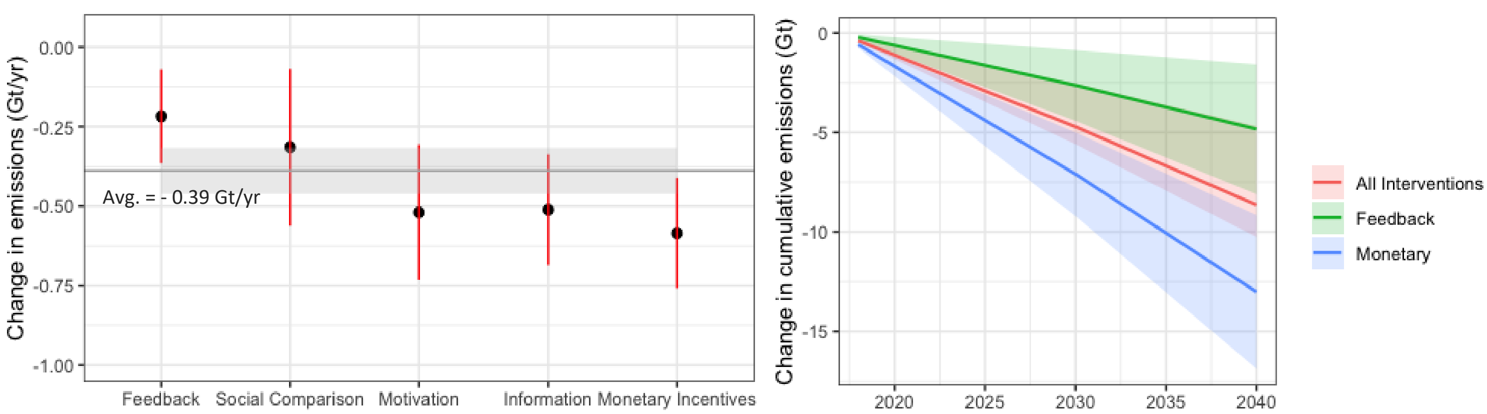
Discussion and outlook

We perform an inter-disciplinary meta-analysis of the effectiveness of pecuniary and behavioral interventions in household energy consumption comprising 122 primary studies and 360 effects sizes representing 1.2 million households in 25 countries. To our knowledge this is the most comprehensive assessment to date. We find a medium-sized, average impact of interventions in household energy consumption. The effect is robust across the meta-analytical models and sub-sets of interventions. The average effect differs by intervention type, with monetary incentives and information being more effective than other interventions—motivation, social comparison, and feedback.

Our findings support the idea that behavioral interventions should not be looked at only individually but rather as packages to increase effectiveness[12,43]. Interventions are usually at least additive and smart packaging can ensure that the overall effect of a portfolio of well integrated interventions is larger than the sum of the separate effects when interventions are applied in isolation. But more research is required to understand why some combinations work better together than others to identify possible trade-offs while combining interventions.

Our moderator variable analysis points towards possibly lower effects for interventions implemented at scale due to self-selection bias, a concern which has also been noted in primary studies[44]. Our analysis also highlights the need for more long term trials, using rigorous methodology and controls for contiguous factors. We are unable to assess persistence of effects after the treatment period[33], which is critical especially for behavioral interventions, but also to an extent monetary incentives. This is because studies do not always include follow up periods and even where they are included, they are not consistent in terms of energy consumption metric, and comparator used (follow up period consumption to treatment period consumption or baseline consumption).

Figure 3 Global average annual (left panel) and cumulative (right panel) CO_2 emissions reduction potential of interventions in household energy demand on building emissions along with the 95% confidence intervals



In spite of these limitations, our meta-analysis offers important insights regarding the carbon emissions mitigation potential of the studied interventions for climate changes assessments. Using percentage reduction in electricity consumption as the dependent variable in our meta-analytical models along with the aggregate emissions of households, we are able to calculate an emissions reduction wedge (see methods). Overall, pecuniary and behavioral interventions in household energy demand can on an average deliver immediate reduction of 0.39 Gt CO_2 yr^{-1} or cumulative reductions of 8.64 Gt CO_2 by 2040 in global carbon emissions of residential buildings (Figure 3). The reduction is higher when only monetary incentives are used and lower when only feedback and social comparison are deployed.

This estimated mitigation wedge is conservative. The reductions could be enhanced by using our evidence on interactions between the various interventions, including the consideration of interaction between injunctive and descriptive norms[45], and the interaction between social norms, behavioral interventions and infrastructure provisions[46] or building design[47]. Cost effectiveness of a basket of interventions should also be assessed by taking into account the costs of different interventions (monetary incentives for example could entail higher infrastructure and regulatory costs). Further, our estimate is based on the current average emissions intensity of electricity grids but would increase if the reductions in energy demand lead to reduction of generation from coal power plants at the margin, as has been the case in the current COVID induced demand reductions[48]. Our estimate also only considers the *reduction* in energy consumption from household interventions but not *shift* in consumption from peak to non-peak hours, which can reduce electricity consumption during peak carbon emissions hours by up to 10%[49]. Finally, our moderator variable analysis does not find significant differences in effectiveness of interventions across regions, and it's reasonable to expect that interventions in energy demand can temper the rapid growth of energy demand in developing countries in South and South-East Asia and sub-Saharan Africa leading to higher savings in emissions. Thus while the estimated carbon mitigation wedge of interventions in residential energy demand is relatively small, the actual impact in specific contexts is likely to be higher. Rightly configured interventions in household energy demand offer a no regret option that can move economies to less risky, low consumption demand pathways towards achieving the Paris climate goals.

References

1. Creutzig, F. *et al.* Beyond Technology: Demand-Side Solutions for Climate Change Mitigation. *Annu. Rev. Environ. Resour.* **41**, 173–198 (2016).
2. Grubler, A. *et al.* A low energy demand scenario for meeting the 1.5 °c target and sustainable development goals without negative emission technologies. *Nat. Energy* (2018) doi:10.1038/s41560-018-0172-6.
3. Van Vuuren, D. P. *et al.* Alternative pathways to the 1.5 °c target reduce the need for negative emission technologies. *Nat. Clim. Chang.* (2018) doi:10.1038/s41558-018-0119-8.
4. Stechow von, C. *et al.* 2 °C and SDGs: united they stand, divided they fall? *Environ. Res. Lett.* **11**, 34022 (2016).
5. IEA. Perspectives for a Clean Energy Transition. The Critical Role of Buildings. *Energy Transit. Prog. Outlook to 2020.* 117 (2019).
6. IEA. *Tracking Buildings 2020*. (2020).
7. Ürge-Vorsatz, D. *et al.* Buildings. In: Mitigation. Working Group III contribution to the Fifth Assessment Report of the Intergovernmental Panel of Climate Change. 671–738 (2014).
8. Masson-Delmotte, V. *et al. 'Global warming of 1.5 C.' An IPCC Special Report on the impacts of global warming*. (2018).
9. Ivanova, D. *et al.* Quantifying the potential for climate change mitigation of consumption options. *Environ. Res. Lett.* **15**, (2020).
10. Hahn, R. & Metcalfe, R. The Impact of Behavioral Science Experiments on Energy Policy. *Econ. Energy Environ. Policy* **5**, 1–23 (2016).
11. Faruqui, A., Arritt, K. & Sergici, S. The impact of advanced metering infrastructure on energy conservation: A case study of two utilities. *Electr. J.* **30**, 56–63 (2017).
12. Stern, P. C. A reexamination on how behavioral interventions can promote household action to limit climate change. *Nat. Commun.* **11**, 10–12 (2020).
13. Faruqui, A. & Sergici, S. Arcturus: International Evidence on Dynamic Pricing. *Electr. J.* **26**, 55–65 (2013).
14. Karlin, B., Zinger, J. F. & Ford, R. The effects of feedback on energy conservation: A meta-analysis. *Psychol. Bull.* **141**, 1205–1227 (2015).
15. Abrahamse, W., Steg, L., Vlek, C. & Rothengatter, T. A review of intervention studies aimed at household energy conservation. *J. Environ. Psychol.* **25**, 273–291 (2005).
16. Andor, M. A. & Fels, K. M. Behavioral Economics and Energy Conservation – A Systematic Review of Non-price Interventions and Their Causal Effects. *Ecol. Econ.* **148**, 178–210 (2018).
17. Nisa, C. F., Bélanger, J. J., Schumpe, B. M. & Faller, D. G. Meta-analysis of randomised controlled trials testing behavioural interventions to promote household action on climate change. *Nat. Commun.* **10**, (2019).
18. Delmas, M. A., Fischlein, M. & Asensio, O. I. Information strategies and energy conservation behavior: A meta-analysis of experimental studies from 1975 to 2012. *Energy Policy* **61**, 729–739 (2013).

19. Collaboration for Enviromental Evidence. Guidelines for systematic review and evidence synthesis in environmental management. *Environ. Evid.* (2013) doi:www.environmentalevidence.org/Documents/Guidelines/Guidelines4.2.pdf.

20. Haddaway, N. R., Macura, B., Whaley, P. & Pullin, A. S. ROSES Reporting standards for Systematic Evidence Syntheses: Pro forma, flow-diagram and descriptive summary of the plan and conduct of environmental systematic reviews and systematic maps. *Environ. Evid.* (2018) doi:10.1186/s13750-018-0121-7.

21. Haddaway, N. R. & Macura, B. The role of reporting standards in producing robust literature reviews. *Nature Climate Change* (2018) doi:10.1038/s41558-018-0180-3.

22. Haddaway, N. R. Open Synthesis: On the need for evidence synthesis to embrace Open Science. *Environ. Evid.* **7**, 4–8 (2018).

23. Sexton, R. J. & Sexton, T. A. Theoretical and Methodological Perspectives on Consumer Response to Electricity information. *J. Consum. Aff.* **21**, 238–257 (1987).

24. Slavin, R. E., Wodarski, J. S. & Blackburn, B. L. 14. 357-363. *J. Aplied Behav. Anal.* **3**, 357–363 (1981).

25. Chen, V. L., Delmas, M. A., Locke, S. L. & Singh, A. Dataset on information strategies for energy conservation: A field experiment in India. *Data Br.* **16**, 713–716 (2018).

26. Gaskell, G. & Pike, R. Residental Energy Use : An Investigation of Consumers and Conservation Strategies. *J. Consum. Policy* **6**, 285–302 (1983).

27. Winett, R. A., Leckliter, I. N., Chinn, D. E., Stahl, B. & Love, S. Q. Effects of Television Modeling on Residential Energy Conservation. *J. Appl. Behav. Anal.* **18**, 33–44 (1985).

28. Winett, R. A., Love, S. Q. & Kidd, C. The Effectiveness of An Energy Specialist and Extension Agents in Promoting Summer Energy Conservation by Home Visits. *J. Environ. Syst.* **12**, 61–70 (1982).

29. Winett, R. A., Neale, M. S. & Grier, H. C. Effects of self-monitoring and feedback on residential electricity consumption. *J. Appl. Behav. Anal.* **12**, 173–184 (1979).

30. Fischer, C. Feedback on household electricity consumption: A tool for saving energy? *Energy Effic.* **1**, 79–104 (2008).

31. Abrahamse, W. & Shwom, R. Domestic energy consumption and climate change mitigation. *Wiley Interdiscip. Rev. Clim. Chang.* **9**, 1–16 (2018).

32. Allcott, H. Social norms and energy conservation. *J. Public Econ.* **95**, 1082–1095 (2011).

33. Allcott, H. & Rogers, T. The short-run and long-run effects of behavioral interventions: Experimental evidence from energy conservation. *Am. Econ. Rev.* **104**, 3003–3037 (2014).

34. Abrahamse, W., Steg, L., Vlek, C. & Rothengatter, T. A review of intervention studies aimed at household energy conservation. *J. Environ. Psychol.* **25**, 273–291 (2005).

35. Ringquist, E. J. *Meta-Analysis for Public Management and Policy*. (Jossey-Bass, 2013).

36. van der Linden, S. & Goldberg, M. H. Alternative meta-analysis of behavioral interventions to promote action on climate change yields different conclusions. *Nat. Commun.* **11**, 10–11 (2020).

37. Funder, D. C. & Ozer, D. J. Evaluating Effect Size in Psychological Research: Sense and Nonsense. *Adv. Methods Pract. Psychol. Sci.* **2**, 156–168 (2019).

38. Abelson, R. P. A variance explanation paradox: When a little is a lot. *Psychol. Bull.* 129–133 (1985).

39. Lanz, B., Wurlod, J. D., Panzone, L. & Swanson, T. The behavioral effect of Pigovian regulation: Evidence from a field experiment. *J. Environ. Econ. Manage.* **87**, 190–205 (2018).

40. Rodemeier, M. & Löschel, A. The Welfare Effects of Persuasion and Taxation: Theory and Evidence from the Field. *SSRN Electron. J.* (2020) doi:10.2139/ssrn.3587339.

41. Viechtbauer, W. Testing Factors and Linear Combinations of Parameters.

42. Srivastava, A., Van Passel, S. & Laes, E. Assessing the success of electricity demand response programs: A meta-analysis. *Energy Res. Soc. Sci.* **40**, 110–117 (2018).

43. Brandon, A., List, J. A., Metcalfe, R. D., Price, M. K. & Rundhammer, F. Testing for crowd out in social nudges: Evidence from a natural field experiment in the market for electricity. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 5293–5298 (2019).

44. Löschel, A. & Werthschulte, M. When Nudges Fail to Scale : Field Experimental Evidence from Goal Setting on Mobile Phones. (2020).

45. Bonan, J., Cattaneo, C., d'Adda, G. & Tavoni, M. The interaction of descriptive and injunctive social norms in promoting energy conservation. *Nat. Energy* **5**, 900–909 (2020).

46. Javaid, A., Creutzig, F. & Bamberg, S. Determinants of low-carbon transport mode adoption: systematic review of reviews. *Environ. Res. Lett.* **15**, (2020).

47. Urge-Vorsatz, D. *et al.* Advances toward a net-zero global building sector. *Annu. Rev. Environ. Resour.* **45**, 227–269 (2020).

48. Bertram, C. *et al.* COVID-induced low power demand and market forces starkly reduce CO2 emissions. *Rev.* (2020).

49. Faruqui, A. & Sergici, S. Household Response to Dynamic Pricing of Electricity - A Survey of the Empirical Evidence. *SSRN Electron. J.* (2011) doi:10.2139/ssrn.1134132.

50. Policies, C. Campbell Collaboration Systematic Reviews: Policies and Guidelines. 1–45 (2014) doi:10.4073/cpg.2016.1.

51. Callaghan, M. W. & Müller-hansen, F. *Statistical Stopping Criteria for Automated Screening in Systematic Reviews*. (2020).

52. Callaghan, M., Müller-Hansen, F., Hilaire, J. & Lee, Y. T. NACSOS: NLP Assisted Classification, Synthesis and Online Screening (Version v0.1.0). (2020) doi:http://doi.org/10.5281/zenodo.4121525.

53. Nelson, J. P. & Kennedy, P. E. The use (and abuse) of meta-analysis in environmental and natural resource economics: An assessment. *Environ. Resour. Econ.* **42**, 345–377 (2009).

54. Viechtbauer, W. Conducting meta-analisys in R with metafor package. *J. Stat. Softw.* **36**, 1–48 (2010).

55. IEA. CO2 Emissions from Fuel Combustion. *2020* https://www.iea.org/reports/co2-emissions-from-fuel-combustion-overview.

Methods

All the information collected in this project is publicly available in line with the systematic reviews reporting protocol (ROSES)[25,26], providing the transparency and reproducibility required to conform with Open Synthesis[22] principles.[22] (see SI for the comprehensive ROSES checklist). We performed a series of meta-analyses on both the full sample as well as (disciplinary) sub-samples in order to assess the effectiveness of different interventions on residential energy consumption. Finally, based on our meta-analyses results we estimate global CO_2 reduction wedge.

Literature search and data extraction: Our data collection strategy involved (1) a search for relevant existing literature reviews and the studies referenced by them; (2) string-based searches of bibliographic databases; and (3) searches for grey literature on Google. In accordance with guidance for rigorous evidence syntheses, we searched a broad set of bibliographic databases (Web of Science Core Collections Citation Indexes, Scopus, JSTOR, MEDLINE), and the web-based academic search engine Google Scholar, based on a comprehensive search string that followed the PICOS (population, intervention, comparator, outcome and study design) logic recommended by the Campbell Collaboration[50]. We developed the search string (see SI) iteratively by checking the results of the search against a set of studies of known relevance. We searched for articles that dealt with household energy (or electricity) consumption along with one or more of interventions of interest. Since we did not make any exclusions based on the date, methodology or the field of publication, the searches returned a large number of studies (64,931) after removing duplicates.

To enable screening of relevant papers, we applied a novel machine learning algorithm using support vector machines[51] to rank the studies in the order of relevance of their abstracts. A team of four reviewers then manually screened the abstracts of the top 6,023 studies. Full text screening was performed on a selection of 939 studies deemed relevant from this initial screening. We only tagged as relevant studies that dealt with energy consumption by households or dormitories and contained a quantitative estimate for the energy saved through a relevant intervention. We did not include studies that focused on price effects but only referenced load effects (changes in kW and not kWh) or those that only reported effect on peak consumption and not total consumption. Studies that only provided an effect size but not the associated variance were not included in the final synthesis. In addition, studies where no obvious comparator group was available (untreated control group or pre-intervention data) or where the sample size was too small to extract meaningful estimates were excluded from the analysis. The final sample included 122 studies after critical appraisal. The inclusion and exclusion criteria, ROSES flowchart for screening and coding and the complete list of studies included in the analysis is available in SI. Four reviewers extracted the relevant data from these studies using the rules laid out in a codebook (see SI). To ensure consistency, a sample of 50 studies was screening at an abstract level (Kappa = 0.77). The reviewers next did a full text screening and coded the relevant papers from this sample, followed by discussion of the coded fields to see what disagreements occurred and suitable adjustments to the codebook. A single reviewer double checked the final data collected for all the included studies. We used the NACSOS software[52] for evidence synthesis developed by MCC Berlin for managing search results, removing duplicates, screening records and extracting data.

Standardizing effect sizes: While the dependant variable in studies in our sample was uniform, relative change in energy consumption, the exact functional form and precision of estimates varied across studies. Since most of the original studies employed regression analysis, following convention[35], we standardized the effects by first converting the regression coefficients extracted from the studies into correlation coefficients r using the total sample size, which were then converted to Fisher's Z . For studies that employed difference of means design, we first calculated the standardised mean differences (smd) or

Cohen's *d* and then converted them to Fisher's Z. The conversions were done using the standard formulae prescribed by Ringquist 2013[35] (see R code in SI for exact conversions).

Synthesis: In order to estimate the aggregate effect size, we first standardized the effects by converting the estimates reported by each study to Fisher's Z[35] We used a random effects model to aggregate the standardized Fisher's Z from the original studies. Random effects model is appropriate when effect sizes in primary studies do not consistently converge to a central population mean[35,53], which is certainly the case for studies relating to energy consumption in households with heterogeneous treatment effects[18]. We used the *metafor* package in R[54] for implementing the random effects model using the DerSimonian-Laird (DL) and restricted maximum likelihood (REML) estimator. Although the DL method is relatively simple and popular, it can lead to severe underestimation of the variance when either the number of studies is limited, or the heterogeneity is large. Instead, Restricted Maximum Likelihood is often recommended, especially when heterogeneity is relatively high[36], so estimating using a REML estimator was preferred. We tested for influential observations using Cook's distance, cov ratio and tau2 (after removal of statistic) diagnostics and identified 8 influential effects. Dropping the influential observations reduced the estimated average effect size to 0.08 – 0.12 but results remained statistically significant and the estimate of tau2 decreased leading to a smaller prediction interval.

Further, even the ordinary random effects model is inappropriate when the effect sizes included are not statistically independent[35]. Effect sizes are likely to dependent in our sample as we extracted multiple effect sizes from each study. In addition, several of the studies in our set employ multiple treatments and some used data from the same underlying experiments. We employed a hierarchical or multilevel meta-analysis model to account for such dependence. The multilevel analysis explicitly models that several of the effect sizes (level 1) come from the same study (level 2). The multilevel analysis used the default variance-covariance structure in the *metafor* package[54]. To test the robustness of our findings we also used cluster robust inference methods using the *clubsandwich* package in R to estimate the variance-covariance matrix (Cluster Robust Variance Estimation). Our results presented in the main paper were robust to the use of these methods.

The meta regression models that investigate the causes for heterogeneity in effect sizes were estimated using REML and multilevel models and introducing moderator variables in the estimation equation. Interaction effects between the various interventions were estimated by including *treatment type* (monetary incentives, information, feedback, social comparison, and motivation) as interacted dummy variables in the estimation equation. The resulting output gives the estimated effect when a single intervention is applied alone and also estimates for all possible combinations of effects seen in the dataset.

Moderator variables for effect heterogeneity: Moderator variables in a meta-regression are factors that influence the conditional expectation of the effect size. Mathematically, the interpretation of the parameter on a moderator variable in meta-regression is the same as for a parameter estimate from a traditional regression; that is, it represents the average change in the effect size associated with one-unit change in the moderator. Moderator variables could represent factors that genuinely affect the magnitude of the relationship between the focal predictor and the outcome of interest or could represent design elements of original studies that may affect effect size from coded studies[35]. In this study we include both type of moderator variables. Design elements of original studies are captured as dummy variables for the following variables: *weather controls* (whether the study controls for it); *demographic controls* (whether the study controls for it); *randomization* and *study design*. The 'other' category of moderator variables captures the factors that are likely to affect the relationship between energy use and the treatment, for example, *duration* of experiment or *region* in which the experiment was performed.

Emissions reductions: To calculate the mitigation wedge, we used the data on direct and indirect CO_2 emissions of households from the IEA[55]. The reduction in electricity consumption was calculated by multiplying the estimated CO_2 emissions of households in 2018 (5.57 Gt) by the average percentage reduction in energy consumption of households due to interventions calculated using the meta-analysis models. The meta-analysis models for this part were run using percentage change in energy consumption reported in primary studies as the dependent variable. The corresponding variance was approximated using square root of sample size[18]. The weighted percentage reduction in energy consumption corresponding with weights from the meta-analysis models was estimated as 6.5% (95% CI = [5.3%, 7.7%]) for the multilevel model. The estimates for cumulative emissions reductions were calculated by assuming the same annual reductions till the respective year.

Data availability: The authors declare that the data supporting the findings of this study are available within the paper and its supplementary information and on Github.

Code availability: All the software packages used for conducting the meta-analysis are open source and freely accessible.