The behavioral mechanisms of habit formation: evidence from a large-scale field experiment

2019 CPMD Workshop on Behavioural and Experimental Economics
UTS, 28 June 2019

Lorenz Goette
University of Bonn
National University of Singapore

Coauthors and partners

The research team

- David Byrne (U Melbourne), Lorenz Goette, Leslie Martin (U Melbourne), Lucy Delahey (SEW), Alana Jones (SEW), Amy Miles (SEW), Samuel Schoeb (U Bamberg), Thorsten Staake (U Bamberg), Verena Tiefenbeck (ETHZ and U Bonn).

▶ Interdisciplinary team from economics, information systems.

The field partners: Technology, financing, and access to participants.

- Southeast Water (SEW), water utility in Melbourne area.
- Amphiro AG

700 households from Melbourne area, part of a larger project by David Byrne and Leslie Martin.

- Recruited to participate from SEW customers.

Outline

Outline

1. Motivation

2. Behavioral mechanisms
   - The consumption-habit model
   - The automatic-control model
   - The attention-habit model

3. Overview of other empirical studies

4. The experimental setup

5. Results
   - Reduced-form evidence
   - Distinguishing behavioral mechanisms

6. Conclusions and implications

Habit formation in economics

Economists' view: habit formation = consumption habit. This is encapsulated in the Stigler and Becker (1977) model:

$$U = u(c_t, h_t) - p_t c_t$$

where habit stock h_t evolves according to an accumulation equation $h_t = \alpha h_{t-1} + (1-\alpha)c_{t-1}$ with $\alpha < 1$.

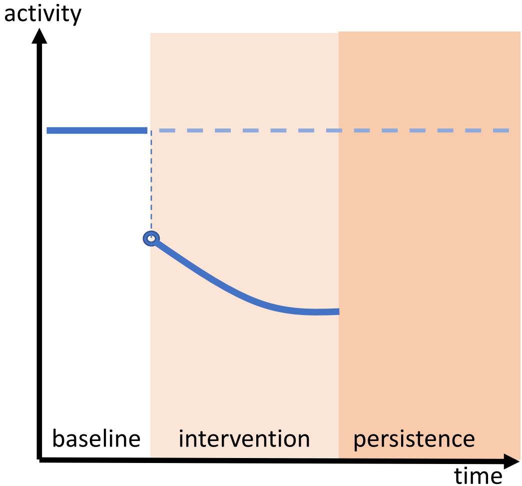
Becker and Stigler (1977): habit formation

A good is habit-forming if

$$\frac{\partial^2 u}{\partial c \partial h} \equiv u_{ch} > 0 \Rightarrow \text{ higher habit stock raises marginal utility of consumption}$$

The implications of consumption-habits for behavior

The canonical habit stock model (Stigler and Becker, 1975): $U = u(a_t, h_t) - \theta p a_t$ where $h_t = \alpha h_{t-1} + (1-\alpha) a_t$ with $u_{ah} > 0$. $\theta < 1$ is salience bias in absence of feedback.



Behavior during the intervention phase.

Intervention phase: Feedback sets $\theta = 1 \Rightarrow$ reduces a (Chetty et al., 2010).

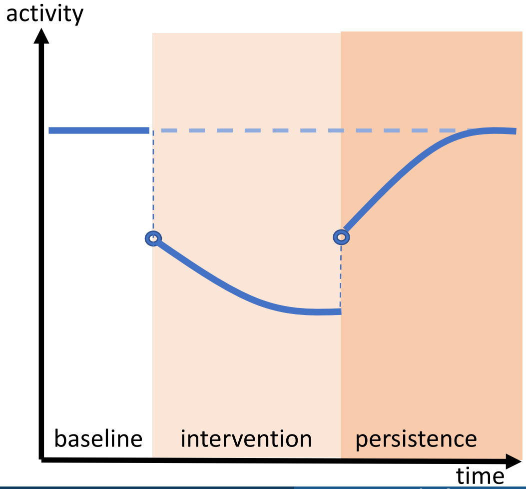
Persistence phase: Reduced activity lowers habit stock

- activity a falls further

Activity level a **converges to new steady state**.

The implications of consumption-habits for behavior

The canonical habit stock model (Stigler and Becker, 1975): $U = u(a_t, h_t) - \theta p a_t$ where $h_t = \alpha h_{t-1} + (1-\alpha)a_t$ with $u_{ah} > 0$. $\theta < 1$ is salience bias in absence of feedback.



Behavior during the persistence phase.

intervention phase: With feedback removed, $\theta < 1$, as before.

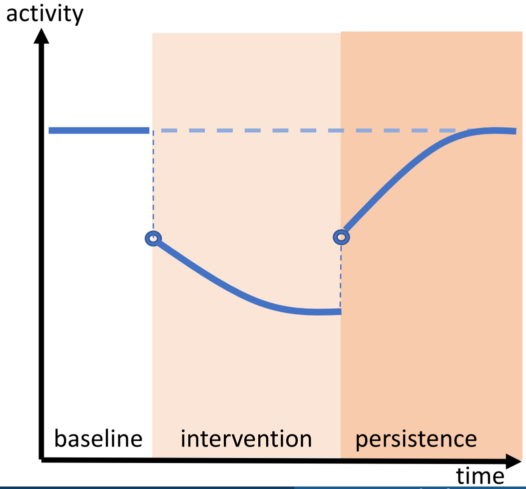
- But due to lower habit stock, activity increases by less.

Persistence phase: Due to increased activity, habit stock rises, and with it consumption.

Activity level **converges back to baseline.**

The implications of consumption-habits for behavior

The canonical habit stock model (Stigler and Becker, 1975): $U = u(a_t, h_t) - \theta p a_t$ where $h_t = \alpha h_{t-1} + (1-\alpha) a_t$ with $u_{ah} > 0$. $\theta < 1$ is salience bias in absence of feedback.



General feature of the model: sluggish response to treatment, gradual reversal to baseline during persistence phase.

The utility function in a automatic-control model

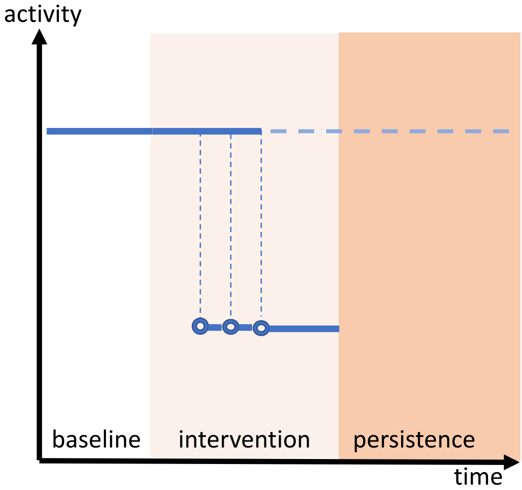
There is none. Psychologists don't write down utility functions (Tsk tsk tsk ...)

View in psychology/neuroscience (Wood and Runger, 2016; Camerer et al., 2017: model coming soon): Habits economize on decision-making.

- Actively evaluating alternatives in each circumstance is cognitively costly.
- It's not optimal to re-optimize every time if the answer is the same every time.

The implications of the shortcut model for behavior

The model (Camerer et al., 2018): individuals stick to a behavior, unless "strong" evidence accumulates that they should re-optimize.



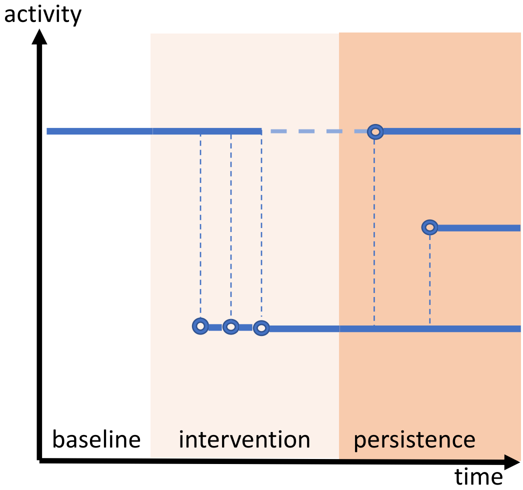
Behavior during the intervention phase.

Individuals respond as enough information accumulates.

Need not be at the same time

The implications of the shortcut model for behavior

The model (Camerer et al., 2018): individuals stick to a behavior, unless "strong" evidence accumulates that they should re-optimize.



Behavior during the persistence phase.

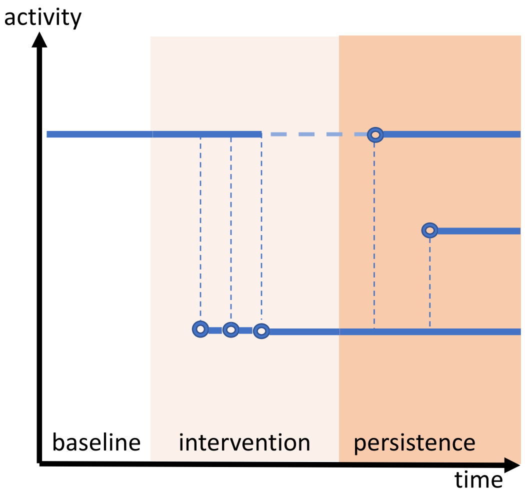
Individuals may switch back to baseline after some time.

Need not be at the same time

Need not switch back to baseline (as information on water use was acquired).

The implications of the shortcut model for behavior

The model (Camerer et al., 2018): individuals stick to a behavior, unless "strong" evidence accumulates that they should re-optimize.



General feature of the model: Discrete jumps by individuals as environment changes (onset or offset of feedback).

Aggregate can looks smooth like in habit stock model.

The utility function in a habit-attention model

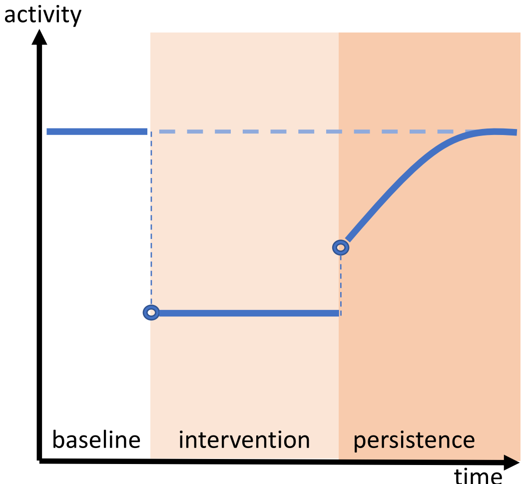
Key idea: attention may be habit-forming. Individuals have limited attention, which leads them to neglect certain aspects of choices (salience bias).

- Sustained attention may create "attention capital": if I have intensely paid attention to attribute A yesterday, this may increase the attention I'm paying to attribute A today.

The implications of attention-habits for behavior.

The model (this paper): $U = u(a_t) - \theta_t p a_t$ where $\theta_t = 1$ with feedback (FB).
$\theta_t = \beta\theta + (1-\beta)\omega_{t-1}$ when FB is off. $\omega_t = \alpha \cdot 1 + (1-\alpha)\omega_{t-1}$ when FB is on.

- Habit formation in attention



Behavior during the persistence phase.

Activity increases, as θ_t drops from 1 to ω_t.

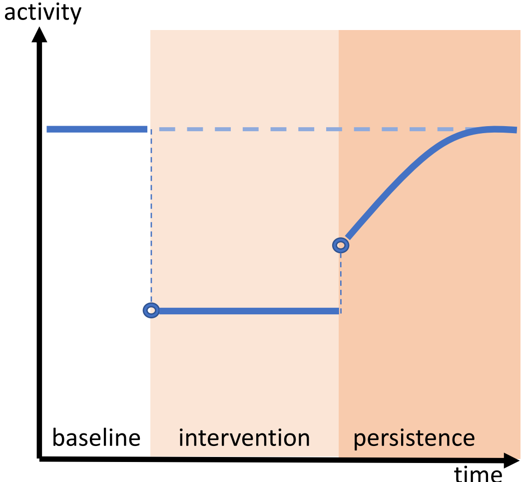
Activity a is below baseline level, as $\omega_t > \theta$.

Gradual erosion as $\theta_t \to \theta$.

The implications of attention-habits for behavior.

The model (this paper): $U = u(a_t) - \theta_t p a_t$ where $\theta_t = 1$ with feedback (FB).
$\theta_t = \beta\theta + (1-\beta)\omega_{t-1}$ when FB is off. $\omega_t = \alpha \cdot 1 + (1-\alpha)\omega_{t-1}$ when FB is on.

- Habit formation in attention

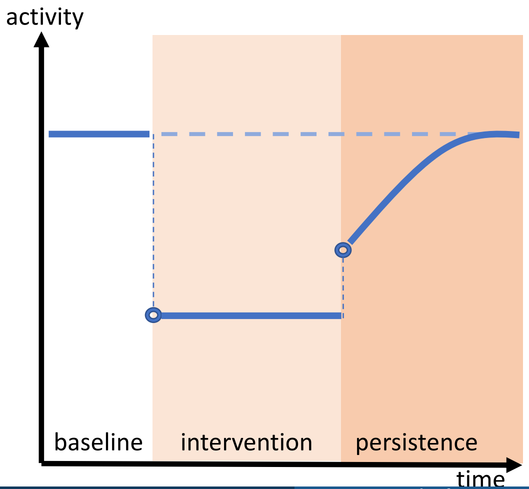


General feature of the model: Discrete, stable drop with onset of feedback. Gradual convergence back to baseline when feedback is off.

The implications of attention-habits for behavior.

The model (this paper): $U = u(a_t) - \theta_t p a_t$ where $\theta_t = 1$ with feedback (FB).
$\theta_t = \beta\theta + (1-\beta)\omega_{t-1}$ when FB is off. $\omega_t = \alpha \cdot 1 + (1-\alpha)\omega_{t-1}$ when FB is on.

- Habit formation in attention

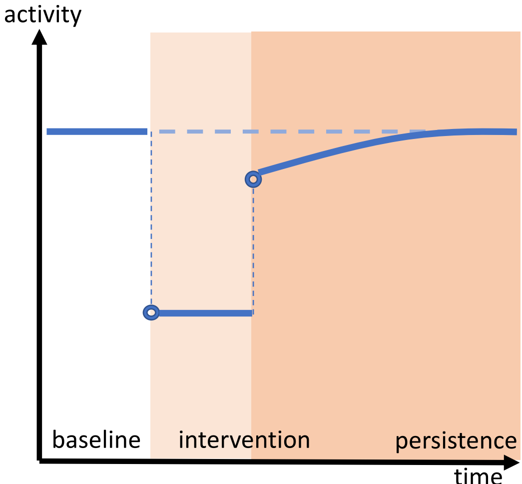


Length of feedback phase determines extent of persistence. Longer feedback phase leads to higher attention to costs ω_t.

The implications of attention-habits for behavior.

The model (this paper): $U = u(a_t) - \theta_t p a_t$ where $\theta_t = 1$ with feedback (FB).
$\theta_t = \beta\theta + (1-\beta)\omega_{t-1}$ when FB is off. $\omega_t = \alpha \cdot 1 + (1-\alpha)\omega_{t-1}$ when FB is on.

- Habit formation in attention

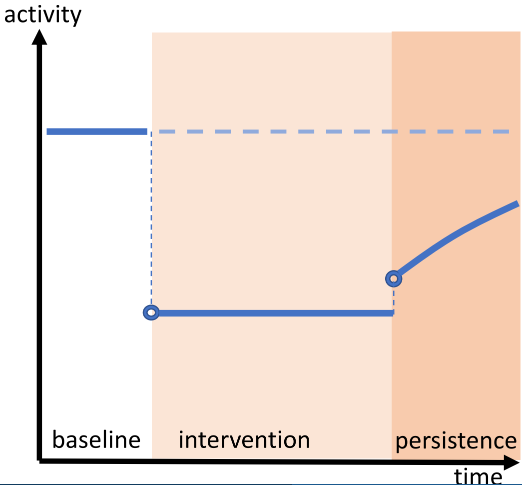


Length of feedback phase determines extent of persistence. Longer feedback phase leads to higher attention to costs ω_t.

The implications of attention-habits for behavior.

The model (this paper): $U = u(a_t) - \theta_t p a_t$ where $\theta_t = 1$ with feedback (FB).
$\theta_t = \beta\theta + (1-\beta)\omega_{t-1}$ when FB is off. $\omega_t = \alpha \cdot 1 + (1-\alpha)\omega_{t-1}$ when FB is on.

- Habit formation in attention

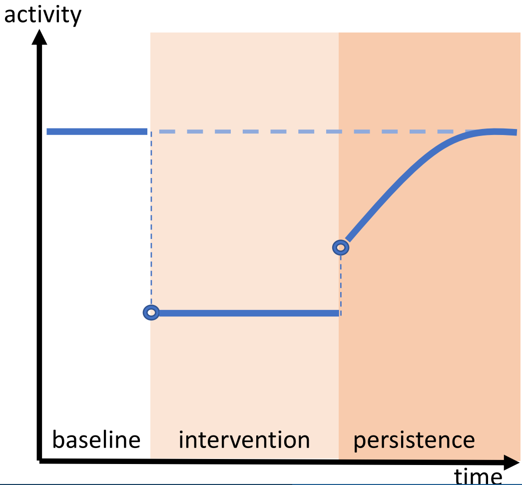


Length of feedback phase determines extent of persistence. Longer feedback phase leads to higher attention to costs ω_t.

The implications of attention-habits for behavior.

The model (this paper): $U = u(a_t) - \theta_t p a_t$ where $\theta_t = 1$ with feedback (FB).
$\theta_t = \beta\theta + (1-\beta)\omega_{t-1}$ when FB is off. $\omega_t = \alpha \cdot 1 + (1-\alpha)\omega_{t-1}$ when FB is on.

- Habit formation in attention



Length of feedback phase determines extent of persistence. Longer feedback phase leads to higher attention to costs ω_t.

Outline

Previous evidence on habit formation

Economics: Typically on/off designs. Stimulus (e.g., incentive) leads individuals to increase some activity, then remove stimulus.

- Does part of the treatment effect persist?
- Mostly "binary" behaviors.

Going to the gym (Charness and Gneezy, 2010; Acland and Levy, 2015); feedback on energy consumption (e.g., Allcott and Rogers, 2014; Ito et al., 2015): confounds change in behavior with investments in technology; hand washing (Hussam et al., 2016); Voting (Gerber et al., 2003, 2010).

Breaking routines: tube strike leads to more efficient routes (Larcom et al., 2015).

Psychology: Few studies with clean identification.

Neuroscience: Mostly animal studies (Baleine and Doherty, 2010; Dolan and Dayan, 2013)

Consensus view in neuro/psychology : habits take at least a month to build up.

Outline

Behavioral impulse: real-time feedback on resource use

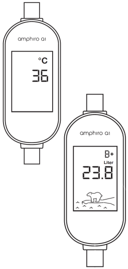
The *amphiro* smart shower meter

Simple and robust design

- Device generates necessary power through water flow
  - ▶ No batteries
- Simple and intuitive display of resource use
  - ▶ Easy to understand
- Saves data (water volume, temperature, duration, breaks) of each shower
  - ▶ Detailed behavioral measures for research use

Behavioral impulse: real-time feedback on resource use

Tiefenbeck et al. (2018), *Management Science*



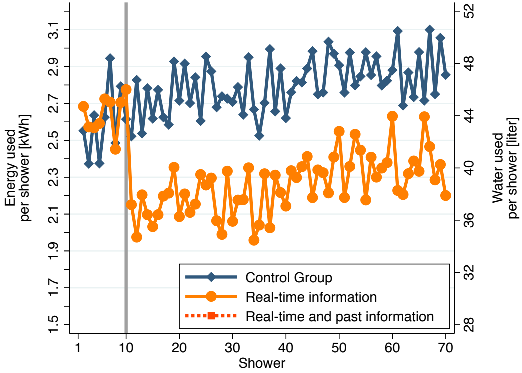
Two experimental conditions:

- Control group: Only sees water temperature.
- Real-time feedback: First ten showers: water temperature. Then real-time feedback.

Behavioral impulse: real-time feedback on resource use

Tiefenbeck et al. (2018), *Management Science*

Comparison of group means over time (study duration: 2 months)



Clear reduction in resource use

- Real-time feedback leads to 22% lower resource use.

- ▶ We use this technology as our behavioral impulse to test for habit formation.

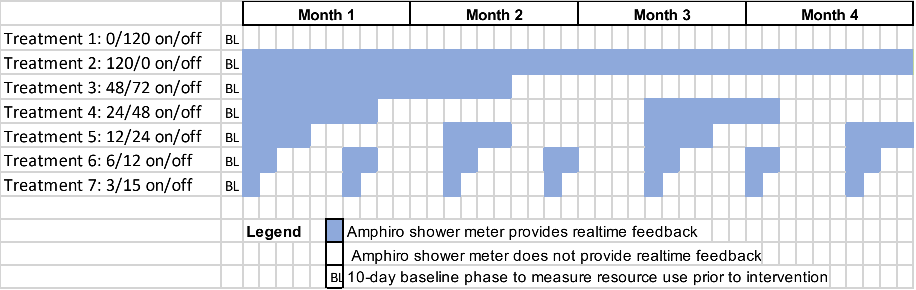
$N = 621$ households, approx $45,000$ showers.

The experimental conditions

Seven experimental conditions

- Control condition: never gets any feedback.
- Real-time feedback condition: gets constant feedback after 20 showers in CG mode.
- Habit-formation conditions: 5 conditions with varying intensity of on-phases and off-phases of feedback after 20 showers in CG mode.

In this design chart, 1 shower per person per day is assumed.



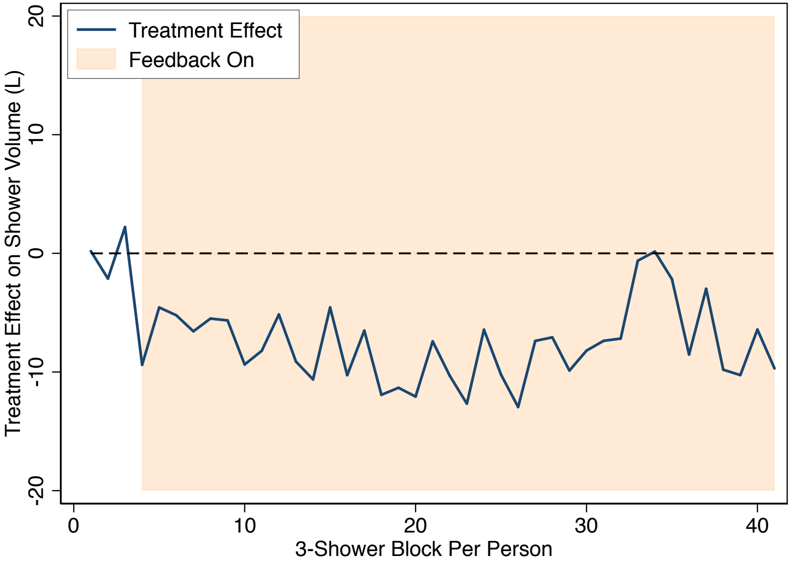
Outline

Descriptive evidence on the treatment effects

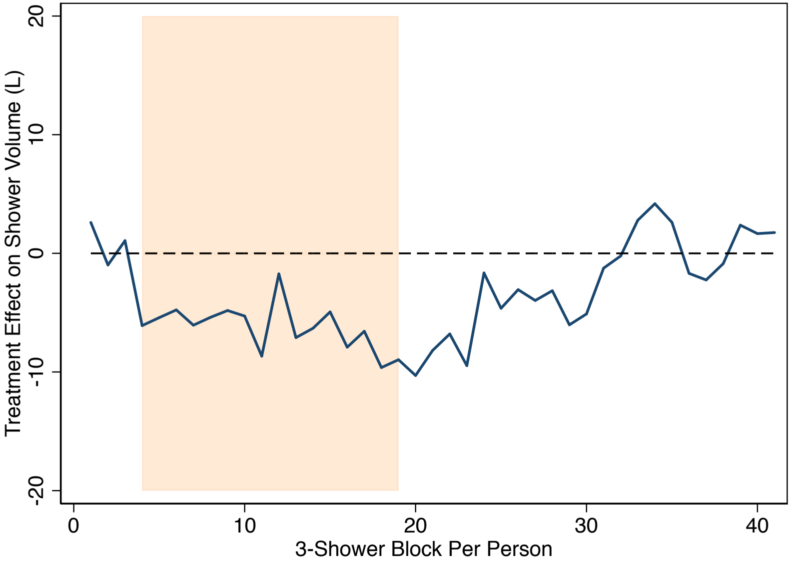
We calculate the average **difference between an experimental condition and the control group** for point in time, subtracting out individual fixed effects.

- Plot these "residualized" averages.

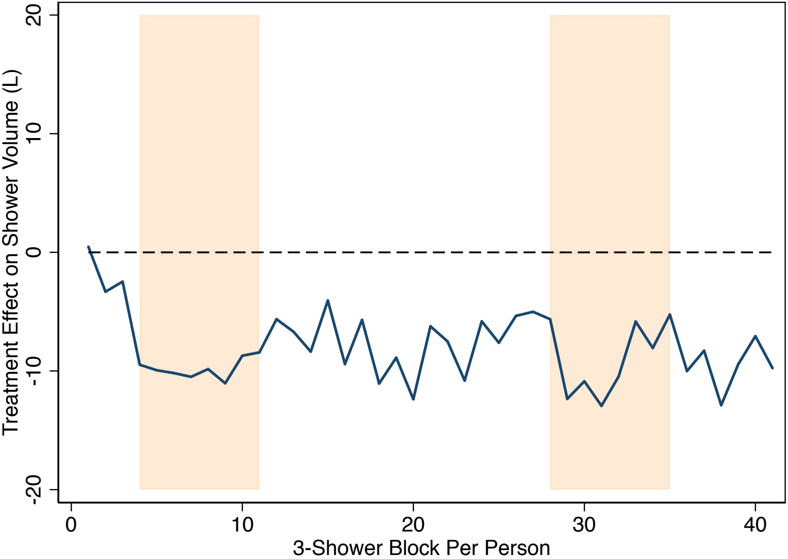
Descriptive evidence on the treatment effects



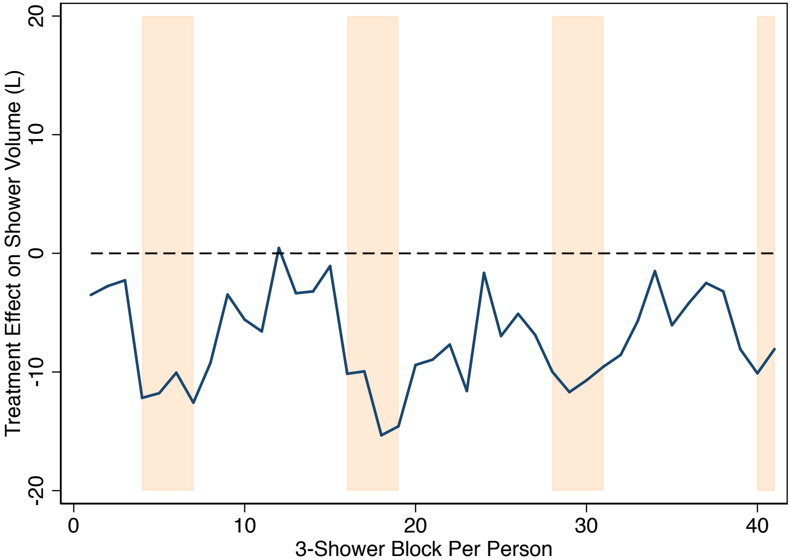
Descriptive evidence on the treatment effects



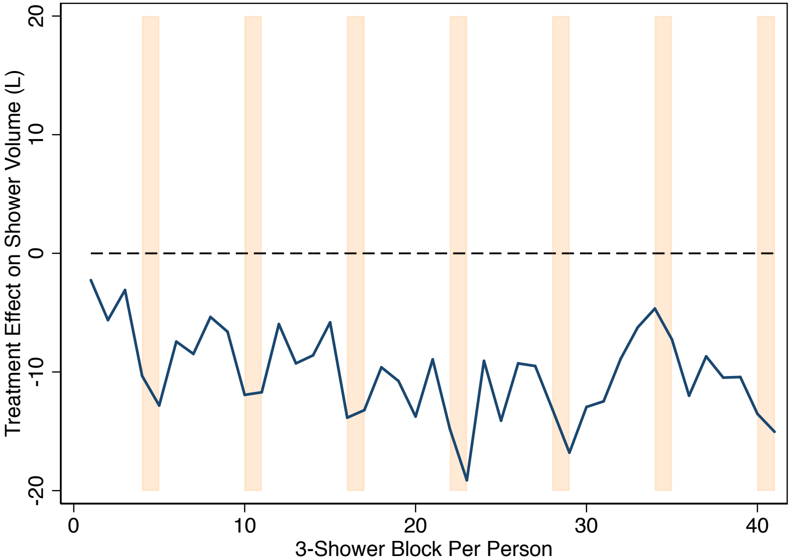
Descriptive evidence on the treatment effects



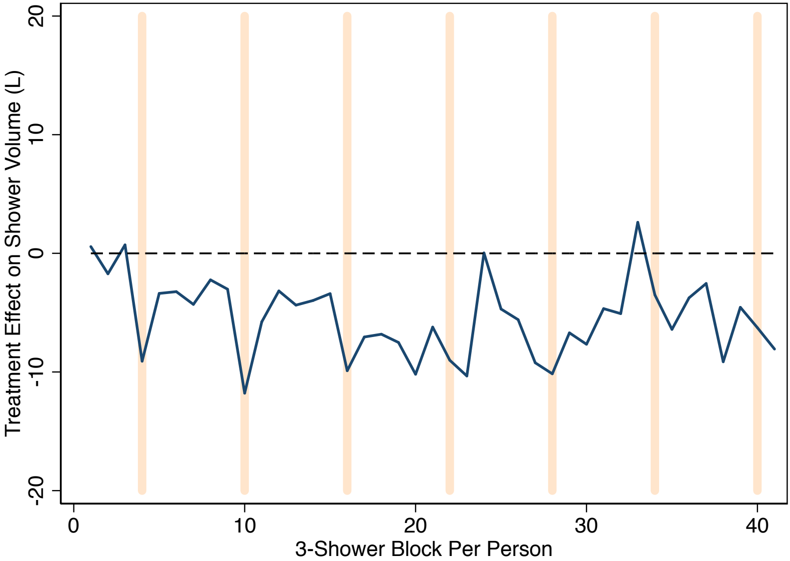
Descriptive evidence on the treatment effects



Descriptive evidence on the treatment effects



Descriptive evidence on the treatment effects



Interpretation

Real-time feedback has a strong effect on use. Replicates findings from previous studies.

A large fraction of the treatment effect persists when feedback is turned off. Evidence of persistence even with ultra-short feedback period (3 showers feedback, 15 showers off).

Difference-in-difference estimates of persistence effects

Estimates for each treatment condition. We estimate the following regression, comparing each of the treatment conditions to the control group and real-time group:

$$y_{is} = \alpha_i + \beta_1 ON_{is} + \beta_2 ON_{is} \cdot PostON_{is} + \beta_3 OFF_{is} + \beta_4 OFF_{is} \cdot PostOFF_{is} + \delta_{t(i,s)} + \epsilon_{is}$$

where

y_{is}	:	water used (in liters) by individual i during shower s.
ON_{is}	:	indicator variable: =1 if feedback is on during shower s, =0 otherwise.
OFF_{is}	:	indicator variable: =1 if feedback had been on before, but is off during shower s, =0 otherwise.
$PostON_{is}$ $PostOFF_{is}$	:	Counts number of showers since feedback was turned on / turned off at shower s.

Difference-in-difference estimates of persistence effects

Estimates for each treatment condition. We estimate the following regression, comparing each of the treatment conditions to the control group and real-time group:

$$y_{is} = \alpha_i + \beta_1 ON_{is} + \beta_2 ON_{is} \cdot PostON_{is} + \beta_3 OFF_{is} + \beta_4 OFF_{is} \cdot PostOFF_{is} + \delta_{t(i,s)} + \epsilon_{is}$$

Interpretation of the estimated coefficients:

α_i	:	Fixed effect for device i
$\delta_{t(i,s)}$	:	Fixed effect for imputed date t.
β_1	:	Impact of feedback on water use.
β_3	:	Impact of previous feedback on water use.
β_2, β_4	:	Change in effectiveness of feedback/previous feedback over time.

Difference-in-difference estimates of persistence effect

Estimates combine control group, real-time group and select persistence conditions H1 - H5, as indicated.

	H1	H2	H3	H4	H5	All
Feedback on (=1)	-8.587**	-7.805**	-7.055**	-8.011**	-7.869**	-7.868**
	(1.399)	(1.339)	(1.537)	(1.415)	(1.474)	(0.917)
FB on $\times$ *PostON*	-0.032	-0.012	-0.023	-0.008	-0.021	0.009
	(0.032)	(0.028)	(0.028)	(0.028)	(0.028)	(0.022)
Feedback Off (=1)	-8.615**	-6.661**	-3.778**	-4.739**	-4.201**	-4.909**
	(1.791)	(1.691)	(1.823)	(1.512)	(1.648)	(0.969)
FB off $\times$ *PostOFF*	0.164**	0.162**	0.283**	0.051	0.061	0.146**
	(0.062)	(0.076)	(0.118)	(0.181)	(0.147)	(0.041)
p-value: no persistence	< 0.001	< 0.001	0.025	0.003	0.019	< 0.001
R^2	0.450	0.475	0.474	0.482	0.473	0.443
Obs	38422	39941	38456	37919	38090	88936

Standard errors clustered on households. ** indicates significance at 5 percent level or better.

Interpretation

Strong evidence of persistence: significant persistence of treatment effects when real-time feedback is turned off.

Persistence forms rapidly. Three showers of feedback are sufficient to induce significant habit formation.

Gradual decay. Our estimates suggest that persistence effects degrade more slowly than they take to build up.

Which is the right behavioral model? Evidence so far leans towards attentional habit model.

Key challenge: testing the automatic-control model

Empirical implications at aggregate level: nearly identical to consumption-habit and attention-habit model.

- In particular, upward drift in off-phase could be generated by either model.

Distinguishing feature of the automatic-control model: individuals change their behavior in discrete jumps rather than smooth trends.

Empirical strategy: allow for individual-level jumps during off-periods.

- Search for the jumps using structural-break tests on each household separately.
- ▶ Given the estimated breaks, do we still find smooth trends during off-periods?

Identifying jumps in post-feedback episodes

We use a **two-step algorithm.**

1. Initializing. For the control group, the real-time group and one of the persistence conditions T3 to T7 and perform the following:

1. Fix one household k in the persistence condition. For an off-phase starting at period t, allow for an additional break at period $t' > t$ for household k

$$y_{is} = \alpha_i + \beta_1 ON_{is} + \beta_2 ON_{is} \cdot PostON_{is} + \beta_3 OFF_{is} + \beta_4 OFF_{is} \cdot PostOFF_{is} + \eta_k \underbrace{OFF_{is} \cdot PostOFF_{is} \cdot \mathbb{1}(i = k) \cdot \mathbb{1}(s \geq t')}_{\equiv PostOFF_{k,s,t'}} + \delta_s + \epsilon_{is}$$

   Record the F-statistic testing the hypothesis $\eta_k = 0$ as $F_{k,t,t'}$.

2. Perform this for all $t' > t$ in the corresponding off-phase. Pick the break t^*_k with the largest F-statistic for this off-period.
3. Perform this for every off-period.
4. Perform this for every individual k in the treatment condition.

Identifying jumps in post-feedback episodes

We use a **two-step algorithm.**

2. Estimation. Having obtained the break dates t_i^* for each individual in each of the persistence conditions, re-estimate the parameters of the entire model:

$$y_{is} = \alpha_i + \beta_1 ON_{is} + \beta_2 ON_{is} \cdot PostON_{is} + \beta_3 OFF_{is} + \beta_4 OFF_{is} \cdot PostOFF_{is} + \eta_k PostOFF_{k,s,t_i^*} + \delta_s + \epsilon_{is}$$

- Re-estimate the model, having assigned "optimal" jumps to all the individuals.

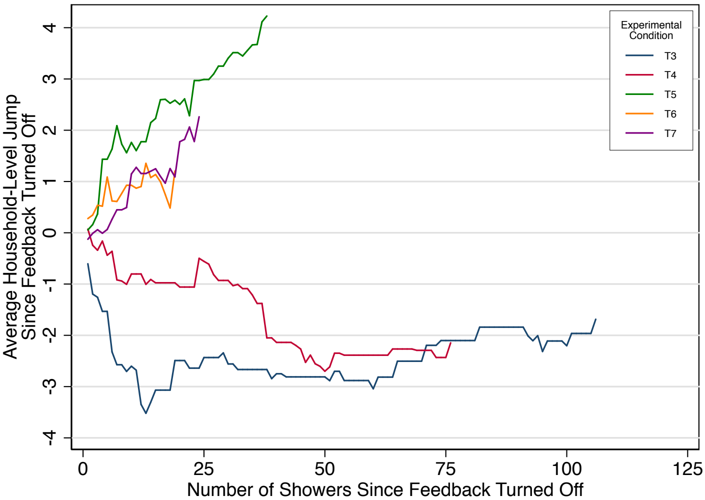
The basic intuition: the algorithm gives discrete jumps the best chance to explain aggregate upward drift during off-phases.

- We estimate the model given these breaks.

- If we still find upward drift $\beta_4 > 0$, this cannot be explained by automatic control mechanisms.

The estimated jumps during off-phases

Implied trend. The jumps are unlikely to explain the drift in off-phases, especially when feedback phase was long.



Persistence estimates, controlling for jumps

Estimates combine control group, real-time group and select persistence conditions H1 - H5, as indicated.

	H1	H2	H3	H4	H5	All
Feedback on (=1)	-8.423**	-7.806**	-7.055**	-8.011**	-7.869**	-7.927**
	(1.401)	(1.345)	(1.537)	(1.415)	(1.474)	(0.927)
FB on $\times$*PostON*	-0.031	-0.015	-0.023	-0.008	-0.021	-0.004
	(0.033)	(0.028)	(0.028)	(0.028)	(0.028)	(0.023)
Feedback Off (=1)	-7.184**	-5.363**	-3.778**	-4.739**	-4.201**	-5.640**
	(1.556)	(1.430)	(1.823)	(1.512)	(1.648)	(0.971)
FB off $\times$*PostOFF*	0.087	0.273**	0.283**	0.051	0.061	0.134**
	(0.056)	(0.070)	(0.118)	(0.181)	(0.147)	(0.036)
p-value: no persistence	< 0.001	< 0.001	0.025	0.003	0.019	< 0.001
R^2	0.456	0.481	0.474	0.482	0.473	0.455
Obs	38422	39941	38456	37919	38090	88936

Standard errors clustered on households. ** indicates significance at 5 percent level or better.

Interpretation of the evidence

Jumps in individual behavior do not explain smooth degradation during off-phases. It happens at the individual level in a continuous fashion even after accounting for jumps.

Results are difficult to reconcile with automatic-control models. The essence of this model is that behavior changes discretely.

Interpretation of the evidence

Jumps in individual behavior do not explain smooth degradation during off-phases. It happens at the individual level in a continuous fashion even after accounting for jumps.

Results are difficult to reconcile with automatic-control models. The essence of this model is that behavior changes discretely.

To be explored: are the mechanisms different between short and long feedback treatments?

Difference-in-difference estimates of persistence effects

Pooled regressions. We estimate the following regression, exploiting variation in habit formation across conditions and time:

$$
\begin{aligned}
y_{is} = \alpha_i &+ \beta_1 ON_{is} + \beta_2 ON_{is} \times PostON_{is} + \beta_3 OFF_{is} + \beta_4 OFF_{is} \times PostOFF_{is} \\
&+ \gamma_1 TotalON_{is} + \gamma_2 OFF_{is} \times TotalON_{is} \\
&+ \eta_k PostOFF_{k,s,t_i^*} + \delta_{t(i,s)} + \epsilon_{is}
\end{aligned}
$$

The specification adds two variables compared to the earlier regressions:

$TotalON_{is}$	:	Variable counting previous exposure to feedback until shower s.
$OFF_{is} \times TotalON_{is}$	:	Interaction term between previous exposure and Off-phase.

Difference-in-difference estimates of persistence effects

Pooled regressions. We estimate the following regression, exploiting variation in habit formation across conditions and time:

$$
\begin{aligned}
y_{is} = \alpha_i &+ \beta_1 ON_{is} + \beta_2 ON_{is} \times PostON_{is} + \beta_3 OFF_{is} + \beta_4 OFF_{is} \times PostOFF_{is} \\
&+ \gamma_1 TotalON_{is} + \gamma_2 OFF_{is} \times TotalON_{is} \\
&+ \eta_k PostOFF_{k,s,t_i^*} + \delta_{t(i,s)} + \epsilon_{is}
\end{aligned}
$$

The regression equation nests the predictions from both models:

- **Consumption-habit model:** $\beta_2 < 0, \gamma_1 < 0$, but $\gamma_2 = 0$.
- ▶ Sluggish response to treatment + previous exposure to feedback leads to stronger response, irrespective of feedback.

- **Attention-habit model:** $\beta_2 = 0, \gamma_1 = 0$, but $\gamma_2 < 0$.
- ▶ Habit-forming exposure only matters when feedback is off.

Difference-in-difference estimates of persistence effects

	Baseline estimates		Estimates controlling for jumps		
Feedback on (=1)	-7.868***	-7.994***	-7.927***	-8.067***	-7.995***
	(0.917)	(0.915)	(0.927)	(0.936)	(0.930)
FB on $\times$ *PostON*	0.009	-0.011	-0.004	-0.040	
	(0.022)	(0.058)	(0.023)	(0.060)	
Feedback Off (=1)	-4.909***	-3.204***	-5.640***	-4.429***	-4.461***
	(0.969)	(1.030)	(0.971)	(1.028)	(1.030)
FB off $\times$ *PostOFF*	0.146***	0.168***	0.134***	0.135***	0.134***
	(0.041)	(0.041)	(0.036)	(0.036)	(0.036)
Total # of FB showers		-0.014		0.025	-0.016
		(0.067)		(0.070)	(0.027)
total FB $\times$ *OFF*		-0.175***		-0.129**	-0.095***
		(0.059)		(0.056)	(0.036)
R^2	0.443	0.443	0.455	0.455	0.455
Obs	88936	88936	88936	88936	88936

Standard errors clustered on households. **, *** indicates significance at 5%, 1% level, respectively.

Outline

Summary of findings

Strong evidence of habit formation at the intensive margin. Our study is the first to show this.

We find striking new patterns in habit formation from feedback. Four key findings:

1. Habits form very rapidly. Three episodes of feedback is enough to build a habit that lasts for 30 episodes.
2. Habits are degrading without feedback: exposure effects slowly wear off.
3. Degrading habit is not explained by heterogeneous jumps (automatic-control interpretation).
4. Intensity of exposure affects behavior when feedback is off. It does not affect behavior when feedback is on.

The behavioral mechanisms underlying habit formation

Our results are inconsistent with the habit-stock model. Key predictions are rejected by the data.

- Habit stock also predicts a sluggish response to the treatment, as habit stock is changed.
- ▶ But impact of feedback on behavior is immediate and stable.

- Previous exposure only affects behavior when feedback is off.
- ▶ Strongly contradicts the notion that habit stock affects marginal utility of activity.

The behavioral mechanisms underlying habit formation

Results constitute first field evidence for attentional habit model. Previous studies tested the implications of the theory only in terms of attention.

- Experiments study reaction times in tasks.

Several features closely parallel findings from attentional habit experiments. Features are inconsistent with habit-stock model:

- Extremely short period of feedback is sufficient to induce persistence.
- Build-up of attentional habit is much faster than decay.

▶ All direct implications from lab studies borne out in our data.

Implications for future research

The impact of the Internet of Things on behavior: Our results show strong habit forming effects from feedback.

- Important question: is this drawing attention from other domains or is it increasing the attention stock overall?
  - Welfare implications differ sharply.

Attentional spillovers to other behaviors? Weak evidence that feedback has positive spillovers to other behaviors (Agarwal et al., 2018; Lade et al., 2017)

- More powerful test in evaluation of PUB's Smart Shower Programme: pilot roll-out of feedback technology to 10,000 households in Singapore.
  - Track water use for showers and total water use to test for spillovers.

That was it.

Thank you for your attention.