Economics & Philosophy (2023), 1–20
doi:10.1017/S026626712300038X

CAMBRIDGE UNIVERSITY PRESS

ARTICLE

The relevance of mechanisms and mechanistic knowledge for behavioural interventions: the case of household energy consumption

Till Grüne-Yanoff[1], Caterina Marchionni[2] and Tatu Nuotio[2]

[1]KTH Royal Institute of Technology, Teknikringen 76, 10044 Stockholm Sweden and [2]University of Helsinki, Unioninkatu 40, 00014 Helsinki, Finland
Corresponding author: Till Grüne-Yanoff; Email: gryne@kth.se

(Received 26 October 2022; revised 20 June 2023; accepted 03 August 2023)

Abstract
We argue that behavioural public policies (BPP) should be categorized by the kind of mechanism through which they operate, not by the kind of treatment they implement. Reviewing the energy consumption BPP literature, we argue (i) that BPPs are currently categorized by treatment; (ii) that treatment-based categories are subject to *mechanistic heterogeneity*: there is substantial variation of mechanisms within each treatment type; and (iii) that they also display *mechanistic overlap*: there is substantial overlap between mechanisms across treatment types. Consequently, current categorizations of BPPs do not reveal the conditions of their efficacy and should be revised to better reflect mechanistic information.

Keywords: Behavioural policy; mechanisms; mediators; moderators; extrapolation

1. Introduction

Mechanistic information is important for assessing the effectiveness and normative acceptability of behavioural public policies (BPP) (see for example, Grüne-Yanoff 2016; Marchionni and Reijula 2019).[1] But how should accepting this argument affect the practices of evaluating the effectiveness of policies? Current practices typically proceeds by experimentally establishing the effect size of a *treatment type* (e.g. default setting, social comparison, self-commitment) and extrapolating the finding to the policy target. In contrast, we argue that treatment categorizations should be supplemented or replaced by *mechanistic categorization*. We develop this argument for behavioural policies aimed at household energy consumption. Nevertheless, as we argue later in the paper, our findings can be generalized to other

[1]We consider behavioural public policies (BPP) to consist of interventions aimed at individual behaviour that are neither coercive nor predominantly operate through financial incentives.

domains of BPP such as recycling and health, where randomized controlled trials (RCTs) and meta-analyses are also common.

Our argument proceeds in three steps. First, we show that current categorizations of behavioural policies in the energy consumption domain are based on the *type of treatment* they implement. The context sensitivity of recorded effect sizes, however, makes it difficult to address questions about extrapolation and ethical evaluation based on these taxonomies alone; mechanistic information is needed too. Second, we argue that supplementing treatment categorizations with mechanistic information, for example by coming up with finer-grained categories, is insufficient when there is both variation of mechanisms *within* each treatment type (mechanistic heterogeneity) as well as overlap between mechanisms *across* treatment types (mechanistic overlap). By surveying treatment types predominant in the literature on energy consumption, we show that there is both substantial mechanistic heterogeneity and mechanistic overlap.[2] We conclude that policy extrapolation and evaluation based on treatment types is hard to justify under such conditions and suggest that current categorizations of behavioural interventions should be revised to better reflect mechanistic information.

The paper is structured as follows. Section 2 describes our notion of mechanism and explains why mechanistic information is important for behavioural policies. Section 3 extracts dominant categorizations from recent reviews of the energy literature and argues that the context sensitivity displayed in this domain puts in doubt that this way of categorizing interventions can handle problems of policy extrapolation and evaluation. Sections 4 and 5 show that there is both substantial mechanistic heterogeneity and overlap in the domain of household energy consumption. Section 6 concludes.

2. Mechanisms and their relevance for behavioural policy

Mechanisms have been characterized as *systems of causally interacting parts and processes, which – under certain conditions – predictably produce one or more effects* (e.g. Glennan 2016; Craver and Tabery 2019). For behavioural policies, mechanisms causally link the policy intervention to the agents' behaviour, making explicit the intermediate stages between them. These intermediate steps, called 'mediators', consist of, for example, components of agents' cognition, social interactions between agents, or tools with which agents take decisions.

Figure 1 represents a mechanism where a *behavioural intervention* consists of a *treatment* (T) that the policymaker implements with the goal of affecting a change in the *behaviour* of individuals (B). Each letter represents a variable, with *mediators* (Me_i) connecting B to T. The directed arrows represent the potential causal influence of one variable on another. A mediator can either pass on causal influence or block it: some variable outside the causal chain might affect a mediator so that it no longer changes in response to the influence of a predecessor in the causal chain (it is 'switched off') or it is less likely to be so affected (its transmission ability is

[2]We draw on the mechanistic explanations proposed in this literature, but we do not evaluate the evidence for these proposals. Instead, we interpret them as potential explanations, which indicate that a certain mechanistic model is considered a genuine possibility by the authors.

https://doi.org/10.1017/S026626712300038X Published online by Cambridge University Press

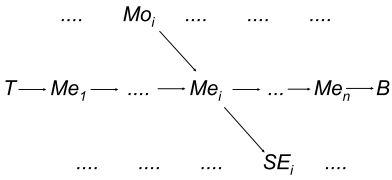

Figure 1. Intervention mechanism.

reduced). Factors that affect a mediator's transmission ability are called *moderators* (Mo_i). In addition to passing on causal influence to their successor in the causal chain, mediators might also have *side effects* (SE_i). Side effects are variables outside the causal chain that are affected by mediators but do not affect them.

Figure 1 is *a* representation of a mechanism. The same mechanism can be represented in different ways. For example, if Me_{i-1} -> Me_i -> Me_{i+1} describes a mechanism, so does Me_j -> Me_{i+1}, where Me_j combines the mediators Me_{i-1} and Me_i as well as their interactions. It is often unknown whether a mediator can be analysed into a sequence of more fine-grained mediators. Furthermore, it is often ontologically uncertain (especially in the social sciences) whether, for a given mechanism, there is an ultimate level of fine-grained mediators. Consequently, we speak of models **M** of a mechanism, where multiple models might be accurate representations of the same mechanism:

> **M**(T,B) is an ordered, non-empty set of variables that represent mediators Me_i, which causally connect T to B, and which also connect to both moderators Mo_i and side effects SE_i.

To illustrate, a policymaker sets a green energy provider default (T). Many consumers see both advantages and disadvantages of this provider in comparison to others, but feel it is too costly for them to form an overall judgement (they would if a more extreme option had been set as default); thus T produces the 'too costly' feeling Me_1 which in turn leads people to stick to the default option (B) (for a discussion of various mechanisms of default-setting interventions, see Grüne-Yanoff 2016). The provision of a simple and trustworthy comparison tool (Mo_1) might reduce this 'too costly' feeling to such an extent that the effect of T on B is blocked. Furthermore, if Me_1 is active, eliciting such a feeling might contribute to a general sense that bureaucratic communications are not worth serious consideration (SE_1).

The philosophical literature offers two arguments for the importance of mechanistic information. First, it helps to make precise the conditions for successful policy extrapolation. An intervention might generate a given outcome in one context (e.g. in an experiment) but fail to do so in another (e.g. in a policy target), because background factors vary greatly between real-world and test conditions and across different contexts. This is why evidence of efficacy from RCTs is generally not sufficient for extrapolation by itself. Instead, justifying inferences of effectiveness from one context to another requires also showing that these contexts do not substantially differ in relevant background factors (Guala 2010; Steel 2010; Grüne-Yanoff 2016; Marchionni and Reijula 2019). How does one know what factors are relevant to justify extrapolation? This is where mechanistic information comes in: if **M** correctly represents the mechanism through which T affects B, then relevant

https://doi.org/10.1017/S026626712300038X Published online by Cambridge University Press

background factors are the moderators Mo_i, which either amplify or inhibit the mediators Me_i of **M**. Mechanistic information thus places the policymaker in a better position to make inferences from the test to the target population: when operating through the mechanism described in **M**, *T* has a chance of success if Mo_i is absent but less so if Mo_i is present.

A second argument for the importance of mechanistic information concerns the normative evaluation of policy applications. A behavioural policy can be assessed either by exclusively focusing on its outcome or by also considering some of its procedural features. A purely outcome-oriented evaluation will need to consider (potentially unintended) side-effects. For example, an intervention that successfully gets residents to lower thermostat settings might still be judged a failure when residents compensate by using portable heating elements, driving up electricity consumption. How does one know that such compensation effects are indeed caused by the intervention rather than, say, by a sudden cold spell? Mechanistic information helps here: by detailing the causal chain from *T* not only to *B* but also to side effects SE_i, knowledge of mechanisms allows the identification and measurement of those variables affected by the intervention, and thus supports a comprehensive evaluation of its outcomes. The importance of mechanisms is even more obvious with respect to evaluations incorporating procedural features. For example, even if an intervention that improves the subjects' knowledge and one that works through social pressure had exactly the same outcome, they might still be evaluated differently. Leveraging social pressure might raise issues of autonomy and subjective well-being if the subjects change their behaviour due to fear of shame or social sanctions. Such procedural information is provided by the mediators Me_i of **M**. Mechanistic information therefore helps evaluating policies from both a purely outcome-oriented and a procedural perspective (Smith *et al.* 2013; Grüne-Yanoff 2016, 2018).

3. Current practice: categorization by treatment

We now show that in the domain of energy consumption behavioural policies have been largely categorized and compared on the basis of the type of treatment they implement. For this purpose, we discuss five popular policy types, as summarized in Table 1.

We claim that these categorizations focus on treatment rather than mechanism for two reasons. First, the terminology in almost all cases only indicates an action of the policymaker or experimenter: they *do* something (set a goal, elicit a commitment, give feedback, etc.) with the purpose of affecting behaviour change. Descriptions of actions do not, however, describe how (in other words, through which mediators) the subjects are affected. To illustrate what we mean here, consider a description that includes such a 'how':

> The social comparison in the message affects behaviour by highlighting social norms rather than by sending signals about privately efficient behaviour (i.e. highlighting cost-savings opportunities). (Ferraro and Miranda 2013: 357)

https://doi.org/10.1017/S026626712300038X Published online by Cambridge University Press

Table 1. Common categorizations of behavioural interventions. To identify the dominant categorizations, we reviewed seven meta-analyses and systematic reviews of multiple behavioural treatments on energy consumption (Row 1). We determined the categorizations used in headings, subheadings and summarizing tables of each article and grouped them based on terminology, similar description in text and literature references (represented in the cells under each article). We labelled each treatment type after the most frequently used term (Column 1). We excluded treatments that do not match the profile of behavioural policies in energy consumption (e.g. material incentives, labelling), and those mentioned in less than two-thirds of the articles

Study type	Abrahamse *et al.* (2005)	Abrahamse and Steg (2013)	Andor and Fels (2018)	Delmas *et al.* (2013)	Khanna *et al.* (2021)	Osbaldiston and Schott (2012)	Šćepanović *et al.* (2017)
Commitment	Commitment	Public commitment	Commitment devices and Goal setting		Commitment devices	Commitment	Commitment, Public commitment
Goal setting	Goal setting		Commitment devices and Goal setting		Goal setting	Goal setting	Goal setting, Group goal setting
Informing	Information			Monetary savings information, Energy saving tips	Information	Instructions	Environmental info, Savings tips
Feedback	Feedback	Group feedback		Individual usage feedback	Feedback	Feedback	Personal feedback
Social comparison	Comparative feedback	Socially comparative feedback, Social norm information and feedback	Social comparison	Social comparison	Social comparison		Comparative feedback, Peer comparison, Social norms

In most cases, the literature only speaks about social comparisons, or similar interventive actions, without specifying how these affect the behaviour and makes such interventive actions the grounds for grouping different studies together. This way of proceeding has the advantage of making it possible to compare effect sizes of different treatments across studies and populations without worrying too much about how the treatment works at the mechanistic level. The assumption is that whatever heterogeneity there is at the mechanistic level, it will be replicated in the target if the effect size is big and robust enough.

Relatedly, the concepts are closely linked to experimental practices: they name treatments implemented in behavioural experiments. These experiments record effect sizes of these treatments, but they do not investigate mechanisms directly. Although the experiments control for some background conditions, they typically do not systematically investigate their influence on effect sizes, nor are there systematic attempts to categorize treatments according to these background conditions. And, as we will discuss in section 4, different types of mediating processes have been proposed as explanations of how the same interventions operate, but usually only informally – without the experiments being designed to test them. Consequently, due to their link with these experimental practices, the categories refer to types of treatment, disregarding procedural or mechanistic aspects.

It could be argued that if effect sizes are robustly associated with treatment types across relevant contexts, this way of categorizing interventions is unproblematic. Unfortunately, empirical evidence contradicts this: the effect size of treatment types are often context-sensitive. To give one example, a recent meta-analysis of feedback interventions aimed at reducing household energy consumption, covering 42 studies, found a mean effect size of 0.071 (Karlin *et al.* 2015). However, the range of individual effect sizes varied from −0.080 to 0.480, i.e. not only did the effectiveness of feedback vary significantly between studies, but several studies even reported negative effect sizes.

What explains this wide range of results? One explanation might be *study quality*, as larger sample size was associated with more modest effects (Karlin *et al.* 2015: 1215). Another might be that the treatment category itself is *too coarse-grained*. For example, effect sizes might differ depending on the specific properties of the feedback provided (Karlin *et al.* 2015: 1220). But such differences can also be due to *context-dependence*, i.e. the influence on the behavioural effect of factors outside the intervention. For example, socially comparative feedback has been found to be more effective for politically liberal individuals than political conservatives in the USA (Costa and Kahn 2013), to have higher impact if subjects are compared with those in their peer network (Peschiera *et al.* 2010) and to have drastically different effects for high- and low-consuming households (e.g. Allcott 2011; Ferraro and Price 2013), with the lowest baseline users even increasing their usage in some cases (e.g. Schultz *et al.* 2007).[3] These contextual factors are examples of mechanistic moderators: they affect the way in which mediators from intervention to behaviour operate

[3]The high context sensitivity of behavioural interventions in the domain of energy consumption has also been noted by Šćepanović *et al.* (2017) and Heiskanen *et al.* (2020). These authors aim to map, for each type of intervention, differences in effect and contextual differences. However, it is only thanks to knowledge of mechanisms that we can reliably identify what contextual factors make a different to the outcome.

https://doi.org/10.1017/S026626712300038X Published online by Cambridge University Press

and, through that, they have an influence on the effect size of the behavioural intervention. Such context dependence of effect sizes exhibits the limitations of trying to answer questions about policy extrapolation and evaluation on the basis of a treatment-based typology alone. At the very least, such categorization would have to be supplemented with mechanistic information. Doing so however is problematic when there is no one-to-one correspondence between treatment and mechanism. If the same treatment operates through different mediators on different occasions, it can also be subject to different moderators and involve side-effects and procedural features that should be evaluated differently. For example, whether social comparison works as optimality signals or through instigating competitiveness may matter to the policymaker beyond its pragmatic utility. The policymaker may be concerned with the ethical consequences of encouraging further competition, for example. More generally, nudges have been criticized for bypassing awareness, suggesting that the way they work matters to policy selection.

4. Mechanistic heterogeneity of treatments

For the five intervention types listed in Table 1, we identify mechanistic hypotheses and explanations proposed in the existing literature and show that there is considerable *mechanistic heterogeneity* in each category.

Recall that we define a mechanistic model **M** as follows: $\mathbf{M}(T,B)$ is a (non-empty) ordered set of mediators Me_i, moderators Mo_i and side effects SE_i, which provides a connection from T to B. Based on this definition, we say that a treatment T is *mechanistically heterogeneous* iff (i) there are multiple models $\mathbf{M_i}(T,B)$, $\mathbf{M_j}(T,B)$ that truthfully represent the causal sequence from T to B, (ii) $\mathbf{M_i}(T,B) \neq \mathbf{M_j}(T,B)$ and (iii) $\mathbf{M_i}(T,B)$ is not a more fine-grained version of $\mathbf{M_j}(T,B)$, – one cannot truthfully replace a causal sequence Me_{i-1} -> Me_i -> Me_{i+1} in model $\mathbf{M_i}$ with a simpler sequence Me_j -> Me_{i+1}, where Me_j combines the mediators Me_{i-1} and Me_i as well as their interactions, yielding $\mathbf{M_i}^{\star}$, such that $\mathbf{M_i}^{\star}(T,B) = \mathbf{M_j}(T,B)$. In other words, there is no truthful representation of the two mechanisms that connect T to B under which they consist of the same mediators, moderators and side-effects. Consequently, if multiple, mutually non-reducible, models are offered as descriptions of the working of the same intervention, we consider it a prima facie reason that the type operates through multiple mechanisms.[4]

4.1. Commitment

A commitment intervention consists in eliciting commitments either to a self-determined or externally imposed goal from subjects who might otherwise not have done so (Katzev and Wang 1994). In a *private pledge*, subjects formulate a pledge only to themselves (Jaeger and Schultz 2017); whereas a *public pledge* is published more widely (Abrahamse *et al.* 2005). In a 2013 meta-analysis, Lokhorst *et al.* (2013) find that during the intervention period reported mean effect sizes during were

[4]In some cases, the multiplicity of models might simply reflect uncertainty about how the treatment works rather than the fact that the treatment works differently under different conditions. Clearly this is a question to be addressed by means of gathering more mechanistic information.

https://doi.org/10.1017/S026626712300038X Published online by Cambridge University Press

medium both for commitment only ($r = 0.27$) and for commitment plus another treatment ($r = 0.31$), but reduced to small in the long run. However, a number of reviews find commitment interventions to be effective in some contexts, for some people and some types of behaviour, but not for others (Osbaldiston and Schott 2012; Frederiks *et al.* 2015; Šćepanović *et al.*, 2017).

In the literature, commitments have been hypothesized to work through the following alternative mechanisms:

(i) Commitment focuses an agent's attention. According to *Feedback Intervention Theory* (Kluger and DeNisi 1996), attention is the drive underlying task performance and must therefore be directed to a specific task-oriented goal. By making subjects pledge to certain goals, the intervention directs their attention to the goals whose realization is being promoted (McCalley 2006).[5] Consequently, actors that affect attention are the most important moderators for commitment interventions that operate through this mechanism.

(ii) Commitment helps agents to create a new goal. According to *Self-Perception Theory* (SPT, Bem 1972), people develop their attitudes by observing their own behaviour and by seeking to rationalize this behaviour by self-attributing these attitudes. Commitment interventions elicit a behaviour (the initial commitment to energy conservation) that, through self-perception, causes the self-attributions of energy-conservation attitudes, which subsequently induce further energy-conseration behaviour. In contrast to (i), the effect of commitment here operates through the creation of new goals, not by directing focus on existing ones. Because it relies on a form of self-rationalization, the most important moderator here is the availability of alternative rationalizations – e.g. coercion, strategic or facile commitment (Burger and Caldwell 2003).

(iii) Commitment changes existing goals. According to *Cognitive Dissonance Theory* (CDT, Festinger 1957), people experience cognitive dissonance when holding contradictory attitudes simultaneously. Commitment interventions might create a new goal that generates such dissonance. To reduce the negative arousal of dissonance experience, people will change attitudes until they become consistent. This might lead to the elimination of previous attitudes and the retainment of the ones to which one is committed to. The CDT mechanism thus differs from the one hypothesized by SPT in at least two respects. First, CDT posits that subjects' original attitudes are clear and important to them, so that they detect the inconsistency, while SPT posits that those original attitudes are relatively ambiguous and less important. Second, CDT posits arousal ('cognitive dissonance') as driving attitude change, while SPT does not (Fazio *et al.* 1977; Schultz 2014). Consequently, ambiguity and alternative sources of

[5]From this perspective, commitment operates through the same mechanism as goal setting. This is probably why some reviewers consider commitment and goal setting as the same intervention type (Andor and Fels 2018).

https://doi.org/10.1017/S026626712300038X Published online by Cambridge University Press

arousal are the most important moderators for commitment interventions that operate through the CDT mechanism.

(iv) Public commitment might affect behaviour by making the subject expect negative social sanctions from non-compliance with the commitment (Cialdini 2001). According to such *social pressure models*, commitment interventions affect subjects' behaviour but not their intrinsic attitudes (Abrahamse *et al.* 2005), which were the driving factors in the previous three cases. Consequently, the subject's perceived publicity of the declaration and the perceived constancy of the witnessing group are the main moderators here – if the subject felt that their promise was no longer public in this sense, the commitment effect would presumably break down. In this view, public commitments work similarly to public social comparison.

4.2. Goal setting

In a goal setting intervention, the goal of saving on energy consumption is formulated explicitly and fixed at a certain level. The goal can be set either by the agents themselves or externally by the experimenter/policymaker. In general, the best performance has been found to take place when the goal is set at a realistic level. In a meta-analysis on goal setting and feedback in general, Neubert (1998) estimates an effect size of 0.63 when feedback was added to goal setting. Similarly, Osbaldiston and Schott (2012) find that goal setting in combination with other treatments has a weighted average effects size ranging from 0.34–1.31. For energy consumption, McCalley and Midden (2002) find that feedback together with goal setting is more effective than feedback alone ($F = 3.6$, $P = 0.03$) with a reduction in energy consumption between 21.9% and 19.5% depending on whether the goal was self-set or assigned.

At least two alternative theories and respective mechanistic hypotheses have been advanced to explain why setting goals can promote behavioural outcomes.

(i) According to Goal-Setting Theory (GST), a specific goal increases performance compared with a vaguely specified goal or no goal (Locke and Latham 1990) through four complementary mechanisms: (a) a goal focuses attention and action toward the desired end and away from irrelevant ones; (b) a goal encourages agents to put an amount of effort proportional to the perceived requirements for attaining the goal (this is why more ambitious goals are *ceteris paribus* more motivating); (c) a goal motivates persistence; (d) a goal activates the skills, knowledge and strategies needed to achieve it, or it motivates people to learn them. For the respective complements, the relationship between goal and performance is claimed to be affected by these moderators: (a) ability, knowledge or skill; (b) commitment to the goal; (c) feedback on goal progress; (d) situational resources. This type of mechanism overlaps with commitment mechanism (i). According to this theory, feedback is a moderator: it motivates persistence and allows the evaluation of current strategies.

https://doi.org/10.1017/S026626712300038X Published online by Cambridge University Press

(ii) An alternative mechanistic hypothesis takes inspiration from *Prospect theory*. The idea is that a specific goal, say reducing energy consumption by 5%, serves as a reference point. Due to framing effects, agents categorize different outcomes as either successes or failures, hence getting closer to one's goal involves positive feelings and vice versa (Heath *et al.* 1999; Harding and Hsiaw 2014). This theory adds a cognitive element to the motivational effect postulated by GST, but in contrast to that, here setting a goal *cannot* change behaviour without a way of knowing how far or close one is to the goal, namely, without some kind of feedback. The most important moderators for goal setting interventions operating through this mechanism are factors affecting the framing.

4.3. Informing

Informing interventions truthfully provide information with the explicit goal of improving people's decision-making. They are distinct from propaganda or deceptive advertisement because they aim to disseminate accurate information only. They are different from education interventions that operate through experiential or affect-based channels – e.g. modelling or gamification – because they spread symbolically encoded information (Dietz and Stern 2002).

A large number of energy conservation experiments use information strategies to reduce energy use (for recent overviews, see Abrahamse *et al.* 2005; Delmas *et al.* 2013). Empirical evidence on the effect of information interventions is mixed. Some researchers conclude that more information has little or no effect on energy use (Abrahamse *et al.* 2005), while others estimate that information programmes could result in energy use reductions on the order of 22–30% over the next 5–8 years (Gardner and Stern 2008). However, these studies examine interventions with different dissemination and information properties, in different contexts, and are therefore difficult to compare. The most recent meta-analysis of informing interventions in this domain found a mean reduction of electricity consumption by 5.61% with a standard deviation of 6.84 (Khanna *et al.* 2021, cf. also Delmas *et al.* 2013, who record a mean effect size of 7.4%). The relatively low effect size, as well as the substantial variations, suggest that any potential causal relationship between information provision and energy consumption is likely mediated by several factors, which in turn are subject to different contextual moderators.

Two alternative mechanistic hypotheses account for the effectiveness of informing interventions.

(i) Information interventions remedy an *information deficit*. The agent lacks information to form justified beliefs about relevant questions and therefore has to deal with uncertainty (Schultz 2002; Darby 2003; Hargreaves *et al.* 2013) or lacks the skills to perform certain actions (e.g. competences to reduce room temperature while remaining comfortable, cf. Corral-Verdugo 2002; Schultz *et al.* 2007). In this case the effectiveness of information interventions depends on three moderators. First, the existence of a relevant *deficit*: if the subjects already possess the information, the intervention will fail. Second, *attention*: without attention, information is not taken up and

https://doi.org/10.1017/S026626712300038X Published online by Cambridge University Press

thus cannot affect behaviour (Clark and Dukas 2003). Third, epistemic *trust* in the medium or source of information: without it, the information will be discounted as evidence and thus will not result in belief expansion or revision (McCraw 2015).

(ii) Even when accurate beliefs and relevant skills are present, information provision can affect behaviour by increasing subjects' motivation. According to *Norm Activation Theory*, for example, changes in behaviour occur when a person is aware of an issue and thinks they can influence it (Vining and Ebreo 2002; Fischer 2008). Informing interventions might raise awareness in norm-activating ways, even when subjects do not lack beliefs or skills. For example, the subjects might already have the information, but the intervention makes the issue salient in a way that motivates them to change their consumption behaviour. In this case, the intervention operates through *focus and attuning* (Kurz *et al.* 2005). Therefore, this mechanism overlaps with commitment mechanism (i) as well as with the first goal-setting mechanism. For this mechanism, attention is an even more important moderator than for deficit-remediation or skill increase, as the subject no longer has independent reason to seek out the information. Informational overload thus is likely to undermine any saliency and thus any effect on behaviour. However, deficit or epistemic trust do not moderate the working of this mechanism.

4.4. Feedback

In feedback interventions, individuals are provided with information about their own energy use that is intended to be more accurate, frequent or accessible than what they currently have. In most cases, feedback includes comparison to past energy usage. Feedback is often combined with other intervention strategies that require individuals to know how much energy they consume, such as social comparison and goal setting.

Two recent meta-analyses found feedback to be an effective intervention strategy in general (Karlin *et al.* 2015; Buckley 2020; but cf. Delmas *et al.* 2013), but they also found results and effect sizes to vary drastically between studies. Some of the differences in effectiveness are likely to be attributable to differences in how feedback is provided (Karlin *et al.* 2015: 1220–1221). Between studies, feedback varies in terms of frequency, medium of delivery, unit of measurement and granularity of information. Accordingly, some meta-analyses do not treat feedback as a homogeneous category. For example, they separate periodical feedback from real-time feedback (e.g. Delmas *et al.* 2013; Buckley 2020).

Feedback interventions can affect behaviour through at least three different pathways:

(i) Feedback *corrects misperceptions* about one's own consumption level. Individuals generally have inaccurate or insufficient knowledge about how much energy they use, which causes them to consume more (or less) energy than they would prefer. Feedback can fill this 'information vacuum' by providing accurate information, enabling households to optimize their energy consumption (Buchanan *et al.* 2014; Matsukawa 2018; Lee *et al.* 2020).

https://doi.org/10.1017/S026626712300038X Published online by Cambridge University Press

In this sense, feedback interventions do not seem to operate differently from informational intervention mechanisms of the first kind and are subject to the same moderators.

(ii) Feedback shows how everyday actions contribute to energy consumption. Feedback provides a link between behaviour and energy consumption and facilitates *learning through experimentation* (Fischer 2008; Buchanan *et al.* 2014; Jessoe and Rapson 2014; Lynham *et al.* 2016; Matsukawa 2018). The capability of feedback to support learning through experimentation is moderated by how close to real-time and how appliance-specific feedback is, so that agents can draw a close enough link between actions and outcomes.

(iii) Feedback serves as a cue that focuses *attention* to energy consumption. Overconsumption results from a lack of awareness when engaging in everyday activities, and feedback reminds individuals to conserve by making the energy consuming nature of activities salient (Fischer 2008; Allcott and Rogers 2014; Carroll *et al.* 2014; Tiefenbeck *et al.* 2018). To the extent that feedback works by directing attention to energy consumption, its effectiveness is likely to depend on *how* it is delivered. Feedback sent to individuals periodically by mail or email may capture attention more easily than feedback provided through smart devices, since the latter requires individuals to engage with it proactively (Karlin *et al.* 2015; Lynham *et al.* 2016). *When* feedback is delivered may also play a role, as providing feedback on a specific energy consuming activity while it is being engaged in may further increase situational focus (Tiefenbeck *et al.* 2018).

4.5. Social comparison

Social comparison interventions provide households with feedback on their own energy consumption *compared with that of others*. Evaluative elements that signal the social desirability of the subject's level of consumption are commonly included. Usually, subjects' consumption is compared with average consumption of *similar* households in the neighbourhood, or with least consuming households. Other reference groups found to be effective include close acquaintances (Peschiera *et al.* 2010) and residents of the same street (Shen *et al.* 2016). By providing information about the behaviour of others, social comparison is a form of *social norm communication*. These types of treatments can communicate either *descriptive norms*, which describe which behaviour is prevalent in a group, or *injunctive norms*, which describe the prevalence of pro-attitudes towards that behaviour in a group (Cialdini *et al.* 1991). In a recent meta-analysis (Khanna *et al.* 2021), social comparison interventions were found to reduce energy consumption by 5.34% on average with a standard deviation of 7.62. Given the complexity of social norms, and the wealth of theories about how they work, it is not surprising that we could identify at least four different mechanisms through which social norms interventions are hypothesized to work.

(i) Information about descriptive norms affects behaviour by being interpreted as *optimality signals*. Observing difference to others' consumption, individuals might conclude that their own consumption habits are suboptimal

https://doi.org/10.1017/S026626712300038X Published online by Cambridge University Press

(Allcott 2011; but cf. Ferraro and Miranda 2013). For example, high consuming individuals might conclude from observing lower average consumption that consuming less is likely to carry some benefit, as suggested by it 'being the norm'. This effect is moderated by the degree to which subjects see the average as establishing such a descriptive norm.

(ii) Comparison with others might motivate some individuals by appealing to the *competitive desire to outdo others* (Abrahamse *et al.* 2005; Andor and Fels 2018). In contrast to the signalling mechanism just described, this competitive mechanism suggests that individuals will seek to move away from the average in whatever direction they deem better. Consequently, the most important moderators for this mechanism are factors affecting competitiveness.

(iii) Social comparison affects the 'moral cost' of energy consumption by eliciting feelings of guilt in those who perceive themselves as violating a social norm and feelings of '*warm glow*' in those who conform (Allcott 2011; Ferraro and Miranda 2013; Delmas and Lessem 2014). For individuals motivated by a concern for social appropriateness, the descriptive norm serves as a reference point that determines socially acceptable levels of energy use (Schultz *et al.* 2007; Andor and Fels 2018). This explanation is supported by the findings that social comparison is more effective on individuals who believe that others care about conserving energy (Jachimowicz *et al.* 2018), and on individuals highly influenceable by what others think they should do (Anderson *et al.* 2017). Consequently, the dominant moderators of this mechanism are those that influence the perception of appropriateness of one's own consumption. For example, communicating descriptive norms to low consuming households might signal to them that it would be reasonable or justified to increase consumption (Schultz *et al.* 2007; Allcott 2011).

(iv) Social comparison, if public, operates through *social pressure*. In most social comparison interventions, individuals' energy consumption is not visible to others, so deviating from the norm has no social consequences. However, when energy consumption is visible to others, individuals can be motivated by the desire to avoid sanctions and maintain a good reputation (Delmas and Lessem 2014). For example, Ferraro and Miranda (2013) found social comparison to have a larger effect on outdoor than indoor water use, which could be explained by outdoor use being more likely to be observed by neighbours. This mechanism overlaps with commitment mechanism (iv). Consequently, factors concerning public observability are the main moderators of this mechanism.

To conclude, we have identified multiple non-reducible mechanistic hypotheses for each of our five selected treatment types. The relevance of the within-type variation was further underscored by identifying different moderators that may amplify or inhibit the effectiveness of the treatment. We summarize our findings in Table 2.

https://doi.org/10.1017/S026626712300038X Published online by Cambridge University Press

Table 2. Mechanistic heterogeneity. Column 1 lists the treatment types identified in Table 1. For each type, column 2 lists mechanistic hypotheses offered in the literature. Column 3 gives examples of potential moderators for the respective mechanisms

Treatment Type	Mechanistic hypotheses	Examples of potential moderators
Commitment	Attention-directing	Attention-moderating factors
	Self-attribution	Alternative rationalizations
	Cognitive dissonance reduction	Ambiguity, other sources of arousal
	Social pressure	Publicity, constancy of witnesses
Goal setting	Motivating effect of setting a specific goal	Ability, knowledge or skill, commitment to the goal, feedback on goal progress, situational resources
	Framing effect from reference point	Framing-moderating factors
Informing	Remedies information deficit	Deficit, attention, trust
	Norm activation	Salience of normative beliefs
Feedback	Correcting misperceptions	Deficit, attention, trust
	Facilitates learning through experimentation	Frequency, appliance specificity
	Attention-directing	Mode of delivery
Social comparison	Optimality signals	Acceptance of average as optimum
	Competition with other consumers	Competitiveness-moderating factors
	Feelings of guilt or 'warm glow'	Concern for social appropriateness of own behaviour
	Social pressure	Publicity

5. Mechanistic overlap

In the previous section, we showed that behavioural policies categorized on the basis of the kind of treatment they implement display substantial mechanistic variation *within* each treatment type, which we refer to as mechanistic heterogeneity. An obvious way to address this heterogeneity is to distinguish further subtypes, identified by the mechanism involved. But simply insisting on more fine-graining often is not an adequate strategy for dealing with BPP categorization. The problem with this strategy is that treatment-based categorizations also display *mechanistic overlap*.

Mechanistic overlap describes a situation where treatments of different types cause their effects through a similar set of mediators. We focus on the simplest case, where distinct treatments produce behaviour through identical mechanisms. Recall that we define a mechanistic model **M** as follows: $\mathbf{M}(T,B)$ is a (non-empty) ordered

https://doi.org/10.1017/S026626712300038X Published online by Cambridge University Press

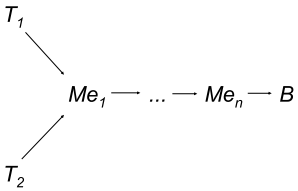

Figure 2. Illustration of mechanistic identity.

set of mediators Me_i, moderators Mo_i and side-effects SE_i, which provides a connection from T to B. Based on this definition, we say that two treatments T_1, T_2 are *mechanistically identical* iff (i) they produce the same B with similar moderator values, and (ii) there are two models $\mathbf{M}(T_1,B)$, $\mathbf{M}(T_2,B)$ such that $\mathbf{M}(T_1,B) = \mathbf{M}(T_2,B)$. Note that there might be many models of the same mechanism for T_1, T_2, respectively. Our definition only requires that *some* models that accurately describe this relation are identical. Mechanistic identity thus is simply the negation of mechanistic difference as defined in section 3. Figure 2 illustrates mechanistic identity so understood.

An example of mechanistic identity is found between public commitment (T_1) and public social comparison (T_2). As discussed above, both might operate through a social pressure mechanism. In response to a public commitment treatment (T_1) individuals identify a group of people taking note of their commitment (Me_1). If group members also follow the commitment, individuals interpret their behaviour as an injunctive norm (Me_2). According to this injunctive norm, they feel pressure to satisfy their own commitment (Me_3), adjusting their behaviour (B) accordingly. When energy consumption of a group of consumers is made public (T_2), consumers identify a group of peers to which they compare themselves (Me_1). If they observe peers sticking with a consumption regime, they interpret their behaviour as an injunctive norm (Me_2). According to this injunctive norm, they feel pressure to avoid sanctions and maintain a good reputation (Me_3), adjusting their behaviour (B) accordingly. Thus T_1 and T_2 operate through the same mechanism under description D.

Note that mechanistic identity does not mean that T_1 and T_2 *always* operate through the same mechanism. As shown in section 4, both interventions might operate through multiple mechanisms. Mechanistic overlap only means that two treatments share one or more mechanisms among those through which they might operate.

Amongst the mechanisms through which each intervention type might operate as described in the previous section, some are identical across types. Table 3 summarizes these mechanistic overlaps.

Different treatment types thus sometimes might operate through the same mechanism. For example, *if* feedback, goal setting and commitment all operate through attention-direction, then they all share the same mechanism and thus require similar moderator values to be instantiated in order to be effective. But they *need not* operate through attention-direction: as we show above, each treatment type can also operate through mechanisms that are not shared between them. Importantly for our argument, the presence of mechanistic overlap shows that not only does treatment type categorization lack detail to discern heterogeneous

https://doi.org/10.1017/S026626712300038X Published online by Cambridge University Press

Table 3. Examples of different treatment types that might operate through identical mechanisms

Mechanism	Treatment types
Attention-directing	Commitment, Goal setting, Feedback
Social pressure	Commitment, Social comparison
Remedies information deficit	Feedback, Informing

mechanisms operating *within* each type, but also fails to account for the same mechanisms operating *across* types, thus disregarding information relevant for extrapolation and evaluation. For example, if an intervention operating through a social pressure mechanism is found to cause negative side-effects under certain conditions, then this could speak against implementing *any* treatment that is likely to operate through the same mechanism in similar contexts. To recognize where extrapolative and evaluative inferences can be generalized across treatment types, a mechanism-based categorization is needed. Mechanistic overlap, therefore, supports the claim that behavioural policies – at least in the energy consumption domain – should be categorized according to mechanisms, not treatment types.

6. Conclusions

In this paper, we accomplished three things: First, we showed that current taxonomies of behavioural interventions aimed at reducing household energy consumption focus on types of treatment rather than on types of mechanisms. Second, we reviewed mechanistic hypotheses about treatment types offered in the literature and showed that interventions of the same type are subject to mechanistic heterogeneity: there is substantial variation of mechanisms *within* each treatment type. Third, we argued that treatment-based categorizations also display mechanistic overlap: there is substantial overlap between mechanisms *across* treatment types.

Although in this paper we only analysed energy consumption BPPs, we believe that our results apply more widely, for two reasons. First, our reading of research papers and reviews from other BPP domains reveals similar categorization practices as documented here. In the literature on health BPPs and on recycling BPPs, we find pretty much the same interventions grouped according to the same criteria as in household energy consumption.[6] Second, many authors in this literature favour RCTs, and consequently focus on measuring effect sizes – i.e. the relation between treatment and behavioural effect. Such a focus generally neglects the investigation of mechanisms, and in particular, it prevents the recognition of mechanistic heterogeneity and overlap.

Our findings weaken the rationale for categorizing behavioural interventions by treatment type. More importantly, they have consequences for the use of treatment

[6]For health BPPs, see reviews by Coupe *et al.* (2019) and Sherrington *et al.* (2016) on weight loss interventions, which discuss commitments and feedback but also mention goal setting and 'information on consequences'. For recycling BPPs, see e.g. Varotto and Spagnolli (2017), which discusses, amongst others, feedback and commitment as intervention types.

https://doi.org/10.1017/S026626712300038X Published online by Cambridge University Press

categorizations for purposes of policy extrapolation. Because of both mechanistic heterogeneity and mechanistic overlap, the conditions of an intervention's efficacy are to be found in the kind of mechanism through which that intervention operates, not in the kind of treatment it implements. This makes extrapolation based on treatment types hard to justify and suggests that current categorizations of behavioural interventions should be revised to better reflect mechanistic information.

References

Abrahamse W. and L. Steg 2013. Social influence approaches to encourage resource conservation: a meta-analysis. *Global Environmental Change* **23**, 1773–1785.

Abrahamse W., L. Steg, C. Vlek and T. Rothengatter 2005. A review of intervention studies aimed at household energy conservation. *Journal of Environmental Psychology* **25**, 273–291.

Allcott H. 2011. Social norms and energy conservation. *Journal of Public Economics* **95**, 1082–1095.

Allcott H. and T. Rogers 2014. The short-run and long-run effects of behavioral interventions: experimental evidence from energy conservation. *American Economic Review* **104**, 3003–3037.

Anderson K., K. Song, S. Lee, E. Krupka, H. Lee and M. Park 2017. Longitudinal analysis of normative energy use feedback on dormitory occupants. *Applied Energy* **189**, 623–639.

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

Bem D.J. 1972. Self-perception theory. *Advances in Experimental Social Psychology* **6**, 1–62.

Buchanan K., R. Russo and B. Anderson 2014. Feeding back about eco-feedback: how do consumers use and respond to energy monitors? *Energy Policy* **73**, 138–146.

Buckley P. 2020. Prices, information and nudges for residential electricity conservation: a meta-analysis. *Ecological Economics* **172**, art. 106635.

Burger J.M. and D.F. Caldwell 2003. The effects of monetary incentives and labeling on the foot-in-the-door effect: evidence for a self-perception process. *Basic and Applied Social Psychology* **25**, 235–241.

Carroll J., S. Lyons and E. Denny 2014. Reducing household electricity demand through smart metering: the role of improved information about energy saving. *Energy Economics* **45**, 234–243.

Cialdini R.B. 2001. *Influence: Science and Practice*. Boston, MA: Allyn & Bacon.

Cialdini R.B., C.A. Kallgren and R.R. Reno 1991. A focus theory of normative conduct: a theoretical refinement and reevaluation of the role of norms in human behavior. *Advances in Experimental Social Psychology* **24**, 201–234.

Clark C.W. and R. Dukas 2003. The behavioral ecology of a cognitive constraint: limited attention. *Behavioral Ecology* **14**, 151–156.

Corral-Verdugo V. 2002. A structural model of proenvironmental competency. *Environment and Behavior* **34**, 531–549.

Costa D.L. and M.E. Kahn 2013. Energy conservation "nudges" and environmentalist ideology: evidence from a randomized residential electricity field experiment. *Journal of the European Economic Association* **11**, 680–702.

Coupe N., S. Peters, S. Rhodes and S. Cotterill 2019. The effect of commitment-making on weight loss and behaviour change in adults with obesity/overweight; a systematic review. *BMC Public Health* **19**, 1–16.

Craver C. and J. Tabery 2019. Mechanisms in science. In *The Stanford Encyclopedia of Philosophy* (Summer 2019 Edition), ed. E.N. Zalta. https://plato.stanford.edu/archives/sum2019/entries/science-mechanisms/.

Darby S. 2003. Making sense of energy advice. In *Time to Turn Down Energy Demand: Energy Intelligent Solutions for Climate, Security and Sustainable Development. eceee Summer Study Proceedings. V. 1–3*, ed. S. Attali, E. Metreau, M. Prone and K. Tillerson.

Delmas M.A. and N. Lessem 2014. Saving power to conserve your reputation? The effectiveness of private versus public information. *Journal of Environmental Economics and Management* **67**, 353–370.

Delmas M.A., M. Fischlein and O.I. Asensio 2013. Information strategies and energy conservation behavior: a meta-analysis of experimental studies from 1975 to 2012. *Energy Policy* **61**, 729–739.

https://doi.org/10.1017/S026626712300038X Published online by Cambridge University Press

Dietz T. and P.C. Stern 2002. Exploring new tools for environmental protection. In *New Tools for Environmental Protection: Education, Information, and Voluntary Measures*, ed. T. Dietz and P.C. Stern. Washington, DC: The National Academies Press.

Fazio R.H., M.P. Zanna and J. Cooper 1977. Dissonance and self-perception: an integrative view of each theory's proper domain of application. *Journal of Experimental Social Psychology* **13**, 464–479.

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

Ferraro P.J. and M.K. Price 2013. Using non-pecuniary strategies to influence behavior: evidence from a large scale field experiment. *The Review of Economics and Statistics* **95**, 64–73.

Festinger L. 1957. *A Theory of Cognitive Dissonance*. Stanford, CA: Stanford University Press.

Fischer C. 2008. Feedback on household electricity consumption: a tool for saving energy? *Energy Efficiency* **1**, 79–104.

Frederiks E.R., K. Stenner and E.V. Hobman 2015. Household energy use: applying behavioural economics to understand consumer decision-making and behaviour. *Renewable and Sustainable Energy Reviews* **41**, 1385–1394.

Gardner G.T. and P.C. Stern 2008. The short list: the most effective actions U.S. households can take to curb climate change. *Environment: Science and Policy for Sustainable Development* **50**, 12–25.

Glennan S. 2016. Mechanisms and mechanical philosophy. In *The Oxford Handbook of Philosophy of Science*, ed. P. Humphreys. Oxford: Oxford University Press. http://works.bepress.com/stuart_glennan/41/.

Grüne-Yanoff T. 2016. Why behavioural policy needs mechanistic evidence. *Economics and Philosophy* **32**, 463–483.

Grüne-Yanoff T. 2018. Boosts vs. nudges from a welfarist perspective. *Revue d'Œconomie politique* **128**, 209–224.

Guala F. 2010. Extrapolation, analogy, and comparative process tracing. *Philosophy of Science* **77**, 1070–1082.

Harding M. and A. Hsiaw 2014. Goal setting and energy conservation. *Journal of Economic Behavior & Organization* **107**, 209–227.

Hargreaves T., M. Nye and J. Burgess 2013. Keeping energy visible? Exploring how householders interact with feedback from smart energy monitors in the longer term. *Energy Policy* **52**, 126–134.

Heath C., R.P. Larrick and G. Wu 1999. Goals as reference points. *Cognitive Psychology* **38**, 79–109.

Heiskanen, E., K. Matschoss, S. Laakso and E-L. Apajalahti 2020. A critical review of energy behaviour change: the influence of context. In *Energy and Behaviour: Towards a Low Carbon Future*, eds M. Lopes, C. Hengelegger Antunes and K.B. Janda, 391–414. Amsterdam: Academic Press.

Jachimowicz J.M., O.P. Hauser, J.D. O'Brien, E. Sherman and A.D. Galinsky 2018. The critical role of second-order normative beliefs in predicting energy conservation. *Nature Human Behaviour* **2**, 757–764.

Jaeger C.M. and P.W. Schultz 2017. Coupling social norms and commitments: testing the underdetected nature of social influence. *Journal of Environmental Psychology* **51**, 199–208.

Jessoe K. and D. Rapson 2014. Knowledge is (less) power: experimental evidence from residential energy use. *American Economic Review* **104**, 1417–1438.

Karlin B., J.F. Zinger and R. Ford 2015. The effects of feedback on energy conservation: a meta-analysis. *Psychological Bulletin* **141**, 1205–1227.

Katzev R.D. and T. Wang 1994. Can commitment change behavior? A case study of environmental actions. *Journal of Social Behavior and Personality* **9**, 13–26.

Khanna T.M., G. Baiocchi, M. Callaghan, F. Creutzig, H. Guias, N.R. Haddaway, L. Hirth, A. Javaid, N. Koch, S. Laukemper, A. Löschel, M. del Mar Zamora Dominguez and J.C. Minx 2021. A multi-country meta-analysis on the role of behavioural change in reducing energy consumption and CO2 emissions in residential buildings. *Nature Energy* **6**, 925–932. https://www.nature.com/articles/s41560-021-00866-x.

Kluger A.N. and A. DeNisi 1996. The effects of feedback interventions on performance: a historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin* **119**, 254–284.

Kurz T., N. Donaghue and I. Walker 2005. Utilizing a social-ecological framework to promote water and energy conservation: a field experiment. *Journal of Applied Social Psychology* **35**, 1281–1300.

Lee, E., M. Kang, J. Song and M. Kang 2020. From intention to action: habits, feedback and optimizing energy consumption in South Korea. *Energy Research & Social Science* **64**, Art. 101430.

Locke E.A. and G.P. Latham 1990. *A Theory of Goal Setting & Task Performance*. Englewood Cliffs, NJ: Prentice-Hall.

Lokhorst A.M., C. Werner, H. Staats, E. van Dijk and J.L. Gale 2013. Commitment and behavior change: a meta-analysis and critical review of commitment-making strategies in environmental research. *Environment and Behavior* **45**, 3–34.

Lynham J., K. Nitta, T. Saijo and N. Tarui 2016. Why does real-time information reduce energy consumption? *Energy Economics* **54**, 173–181.

Marchionni C. and S. Reijula 2019. What is mechanistic evidence, and why do we need it for evidence-based policy? *Studies in History and Philosophy of Science* **73**, 54–63.

Matsukawa I. 2018. Information acquisition and residential electricity consumption: evidence from a field experiment. *Resource and Energy Economics* **53**, 1–19.

McCalley L.T. 2006. From motivation and cognition theories to everyday applications and back again: the case of product-integrated information and feedback. *Energy Policy* **34**, 129–137.

McCalley L.T. and C.J.H. Midden 2002. Energy conservation through product-integrated feedback: the roles of goal-setting and social orientation. *Journal of Economic Psychology* **23**, 589–603.

McCraw B.W. 2015. The nature of epistemic trust. *Social Epistemology* **29**, 413–430.

Neubert M.J. 1998. The value of feedback and goal setting over goal setting alone and potential moderators of this effect: a meta-analysis. *Human Performance* **11**, 321–335.

Osbaldiston R. and J.P. Schott 2012. Environmental sustainability and behavioral science: meta-analysis of proenvironmental behavior experiments. *Environment and Behavior* **44**, 257–299.

Peschiera G., J.E. Taylor and J.A. Siegel 2010. Response–relapse patterns of building occupant electricity consumption following exposure to personal, contextualized and occupant peer network utilization data. *Energy and Buildings* **42**, 1329–1336.

Schultz P.W. 2002. Knowledge, information, and household recycling: examining the knowledge-deficit model of behavior change. In *New Tools for Environmental Protection: Education, Information, and Voluntary Measures*, eds T. Dietz and P.C. Stern, 67–82. Washington, DC: The National Academies Press.

Schultz P.W. 2014. Strategies for promoting proenvironmental behavior: lots of tools but few instructions. *European Psychologist* **19**, 107–117.

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

Šćepanović S., M. Warnier and J.K. Nurminen 2017. The role of context in residential energy interventions: a meta review. *Renewable and Sustainable Energy Reviews* **77**, 1146–1168.

Shen M., R. Young and Q. Cui 2016. The normative feedback approach for energy conservation behavior in the military community. *Energy Policy* **98**, 19–32. ISSN 0301-4215. https://doi.org/10.1016/j.enpol.2016.08.014.

Sherrington A., J.J. Newham, R. Bell, A. Adamson, E. McColl and V. Araujo-Soares 2016. Systematic review and meta-analysis of internet-delivered interventions providing personalized feedback for weight loss in overweight and obese adults. *Obesity Reviews* **17**, 541–551.

Smith N.C., D.G. Goldstein and E.J. Johnson 2013. Choice without awareness: ethical and policy implications of defaults. *Journal of Public Policy & Marketing* **32**, 159–172.

Steel D. 2010. A new approach to argument by analogy: extrapolation and chain graphs. *Philosophy of Science* **77**, 1058–1069.

Tiefenbeck V., L. Goette, K. Degen, V. Tasic, E. Fleisch, R. Lalive and T. Staake 2018. Overcoming salience bias: how real-time feedback fosters resource conservation. *Management Science* **64**, 1458–1476.

Varotto A. and A. Spagnolli 2017. Psychological strategies to promote household recycling: a systematic review with meta-analysis of validated field interventions. *Journal of Environmental Psychology* **51**, 168–188.

Vining J. and A. Ebreo 2002. Emerging theoretical and methodological perspectives on conservation behavior. In *Handbook of Environmental Psychology*, eds R. Bechtel and A. Churchman, 541–558. New York, NY: Wiley.

Till Grüne-Yanoff is Professor of Philosophy at the Royal Institute of Technology (KTH) in Stockholm. His research focuses on the philosophy of science, the philosophy of economics and on decision theory. In particular, he investigates the practice of modelling in science and engineering, develops formal models of

https://doi.org/10.1017/S026626712300038X Published online by Cambridge University Press

preference consistency and preference change and discusses the evaluation of evidence in policy decision making.

Caterina Marchionni is a researcher in Practical Philosophy at the University of Helsinki and member of TINT – Centre for Philosophy of Social Sciences. Her primary research areas are the philosophy of economics and the philosophy of the social and human sciences. In her work, Caterina has been paying special attention to issues of explanation, models, and evidence across disciplinary boundaries. Email: caterina.marchionni@helsinki.fi

Tatu Nuotio is a MSSc from the University of Helsinki and affiliated with TINT – Centre for Philosophy of Social Science. His research interests centre on the topics of social norms and norm-related policy. Email: tatu.nuotio@gmail.com

Cite this article: Grüne-Yanoff T, Marchionni C, and Nuotio T. The relevance of mechanisms and mechanistic knowledge for behavioural interventions: the case of household energy consumption. *Economics and Philosophy*. https://doi.org/10.1017/S026626712300038X