Testing for the size heuristic in householders' perceptions of energy consumption

Laura Cowen, Birgitta Gatersleben

School of Psychology, University of Surrey, Guildford, GU2 7XH, United Kingdom

Author Note

Email: l.cowen@surrey.ac.uk

Abstract

Few householders have the time or motivation to systematically weigh up all the facts when judging the energy consumption of their household appliances. It is likely that they instead rely on simple heuristics such as the size heuristic, which has been reported in a small number of previous studies. The studies showed that people's perceptions of the size and energy consumption of appliances were positively correlated but the studies differed in their methods and effect sizes. The present study re-tests the use of the size heuristic using two methods of data collection (between-participants and within-participants) and three methods of correlation. On average, correlations between size and energy estimates were moderately strong but they (and the accuracy of the energy estimates) varied greatly between individual participants. Understanding householders' perceptions of energy is vital to designing more effective energy-saving policies. The findings highlight the importance of choosing and clearly reporting methods.

Keywords: energy consumption, size heuristic, perception, correlation, methods

Testing for the size heuristic in householders' perceptions of energy consumption

1 Introduction

People are given increasing amounts of information and feedback about their household energy use but a long tradition of decision-making research shows that people do not, and often cannot, seek out, take in, and systematically weigh up all of the relevant facts before making a judgement. Instead, they use *heuristics*, rules-of-thumb, to simplify the complex and numerous judgements they have to make in their everyday lives. It is important to understand then what heuristics people intuitively use to make judgements about energy consumption. Understanding people's mental models of energy use is essential to being able to design effective behaviour-changing interventions and policies in order to reduce household energy consumption.

A small number of studies over the past 35 years have found evidence to suggest that people use a type of heuristic known as *attribute substitution* to help them infer how much energy is consumed by individual appliances in their households:

“when confronted with a difficult question people often answer an easier one instead, usually without being aware of the substitution.” (Kahneman & Frederick, 2002, p. 53)

The studies used a range of methodological approaches and focused on revealing some of the attributes (such as size, duration of use, visibility) of energy-consuming appliances that householders might use as heuristics (the easier question) when judging their energy consumption (the difficult question) (Baird & Brier, 1981; Chisik, 2011; Kempton & Montgomery, 1982; Schley & DeKay, 2015; Schuitema & Steg, 2005; van den Broek, 2016). The most reported attribute, in both qualitative and quantitative studies, is size (Baird & Brier, 1981; Chisik, 2011; Schuitema & Steg, 2005; van den Broek, 2016). That is, when people are judging the amount of energy consumed by an appliance, they appear to actually be judging the size of the appliance: they perceive large appliances to consume large amounts of energy and small appliances to consume small amounts. Two sets of quantitative studies found strong positive correlations between participants' perceptions of the size of appliances and their perceptions of how

much energy the appliances consumed. The correlation coefficient (the effect size) in one study (Schuitema & Steg, 2005, $r = .67$), however, was significantly smaller than in the others (Baird & Brier, 1981, Exp. 2: $r_s = .91$, Exp. 3: $r_s = .93$), so it is not clear how strongly people associate the size and energy consumption of appliances. The two sets of studies used different methods to correlate their data and neither used the between-participants design recommended in the heuristics literature to ensure the relationship between estimates is not a consequence of participants providing both estimates (Kahneman & Frederick, 2002; Kahneman & Tversky, 1996; Tversky & Kahneman, 1983).

In this paper, we re-tested the association between perceived size and perceived energy consumption by using both a between-participants and a within-participants design to test whether the correlation between size and energy consumption estimates still occurs when there is no possibility of priming or other influences typically associated with within-participants study designs. We then used three different methods of correlation on the within-participants dataset to investigate whether different methods of correlation produce different results, like those of Baird and Brier (1981) and Schuitema and Steg (2005). It is important to study how people make judgements about energy consumption so that householders' mental models of energy use are better understood and energy-saving policies can be designed more effectively. It is also important to understand how the different methods that researchers use to conduct such studies can influence their results and, potentially, their conclusions about householders' mental models.

1.1 Perceptions of Household Energy Consumption and Saving

International goals to reduce CO₂ emissions levels include reducing domestic energy consumption (e.g., European Union, 2017a, 2017c). Households and passenger transport together account for 55% of total energy use in the UK (DECC, 2016), approximately 42% in the European Union (European Union, 2017b; US Energy Information Administration, 2015), and approximately 38% in the US (US Energy

Information Administration, 2015, 2016). Although efficiencies in household appliances are currently offsetting the increasing numbers of electricity-consuming devices in UK households, in the US, the reverse is true (DECC, 2016; US Energy Information Administration, 2013).

It is often assumed that people lack accurate knowledge about energy saving and that providing the missing information will help them to change their behaviours and save more energy. Despite the popularity of this approach, there is significant evidence that education alone leads to little or no energy savings (Burgess, Harrison, & Filius, 1998; Delmas, Fischlein, & Asensio, 2013; Kollmuss & Agyeman, 2002). One reason for this might be that there is limited understanding of people's perceptions of energy. The information deficit model implies that information provision would result in more accurate mental models, and more accurate mental models would improve decision making. This may not always be the case, however.

For example, Kempton (1986) interviewed householders about their perceptions of how central heating thermostats work. He showed that the inaccurate mental model held by a large proportion of US householders would be more effective in saving energy than the mental model that heating experts would be more likely to consider correct. This is because the more accurate mental model leads to efficient behaviour only if the householder understands the model almost to the level of a heating expert, which most will not take the time to do. The simpler model, though technically inaccurate, was estimated to lead to behaviours that consume less energy.

More recently, Revell and Stanton (2014) showed that householders might hold various mental models of central heating systems that vary greatly from how the systems actually work. As Norman (1983) pointed out, it is not necessary for people to have accurate mental models of a device or system in order to use it but they must have models that produce the result they want. That is, people don't need to understand the technical details of how the central heating system works as long as they feel comfortable without spending a lot of money. That can be aided by designing a thermostat that appears to match the perceptions of its users, instead of taking the

information deficit approach of trying to change users' perceptions of a complex device. It might not always be relevant to re-design a single device to ensure that people's mental models of energy consumption and saving are effective but the same principle could be applied to the design of energy-saving policies and interventions.

Our approach, therefore, is to develop an understanding of householders' mental models of energy consumption and saving. We are building on a small amount of previous research, some of which has focused on identifying the heuristics that householders use when making judgements about how much energy appliances consume. We observed, however, that there are inconsistencies in methodology that could influence research findings and conclusions. In the study reported in this paper, we investigated the methods that have previously been used, and could be used in future studies, to identify heuristics that people use to judge energy consumption. It is possible that people's mental models of energy saving are different from their models of energy consumption (Attari, DeKay, Davidson, & Bruine de Bruin, 2010, S.I. Text). If reliable methods can be established for investigating people's perceptions of energy consumption based on existing research, future studies can then use them to investigate people's perceptions of energy saving and, also, whether their perceptions are related to actual energy-saving behaviours.

1.2 Using Heuristics to Judge Household Energy Consumption

Mains-supplied electricity and gas are largely invisible to end-users (Burgess & Nye, 2008; Kidd & Williams, 2008) and cannot be directly seen, touched, or heard. Householders can only infer indirectly how much is consumed by their individual appliances so it is plausible that they might base their inferences on the physical attributes—for example, the size—of the appliance itself.

Evidence that people use a size heuristic in their energy consumption judgements comes from a small number of quantitative and qualitative studies (Baird & Brier, 1981; Chisik, 2011; Schuitema & Steg, 2005; van den Broek, 2016). When asked to draw the five highest-energy-consuming things in their home, Chisik's interviewees tended to

draw large appliances and mentioned size and duration of use as reasons for their greater energy consumption. During focus group discussions in which participants collaboratively ranked appliances according to their energy consumption, van den Broek recorded a range of potential heuristics mentioned by the participants, including size. Baird and Brier (1981) and Schuitema and Steg (2005) both asked each of their participants to estimate the size and energy consumption of a set of household appliances and they then correlated the estimates. They found positive correlations between the size and energy consumption estimates, which suggests that people might have been judging energy consumption based on the size of the appliances. Additional support for this conclusion came from their observations of errors made by their participants. Participants overestimated the energy consumption of some large appliances that actually consumed relatively little energy, like the colour TV and stereo (Baird & Brier, 1981, Experiments 2 & 3) or the spin-dryer and electric underfloor bathroom heating (Schuitema & Steg, 2005). They also underestimated the energy consumption of some small appliances that actually consumed relatively large amounts of energy, like the carving knife, hair dryer, and toaster (Baird & Brier, 1981, Experiments 2 & 3) or the coffee maker, VCR, and satellite receiver (Schuitema & Steg, 2005). They appeared to be substituting the question about energy consumption with a question about size, which would normally work as a heuristic but failed for appliances that did not follow that pattern.

Although Baird and Brier (1981, Experiments 2 & 3) and Schuitema and Steg (2005) appeared to be testing for attribute substitution (the substitution of size for the energy consumption of the appliances), their studies did not use the *heuristic elicitation design* method recommended in the heuristics literature (Kahneman & Frederick, 2002; Kahneman & Tversky, 1996) to be most appropriate for identifying this type of heuristic. Their methods of analysis also varied, which might explain why they obtained quite different coefficients. It is important to better understand whether the study designs influenced the results before further research is done into householders' perceptions of energy consumption so that future study findings can be easily

interpreted and compared.

1.3 Testing Heuristics With the Heuristic Elicitation Design

The heuristic elicitation design is a simple method in which the estimates of one group of participants on one variable are correlated with the estimates of another group of participants on another variable (Kahneman & Frederick, 2002). A strong correlation is interpreted as indicating that the participants of one group are actually answering the question asked of the other group, possibly without even realising that they are substituting the questions. The heuristic elicitation design method was used in some of the classic heuristics studies (e.g. Kahneman & Tversky, 1973, 1982) as well as some more recent ones (e.g. Kahneman, Schkade, & Sunstein, 1998; Monin, 2003). A defining feature of the heuristic elicitation design method is that it uses a between-participants design; Baird and Brier (1981, Experiments 2 & 3) and Schuitema and Steg (2005) both used a within-participants design to collect their data.

1.3.1 Within-participants versus between-participants designs. The heuristic elicitation design method uses a between-participants design for data collection so that participants' estimates are not influenced by either their previous estimates or by clues to what the study might be about (Greenwald, 1976; Kahneman & Frederick, 2002; Kahneman & Tversky, 1996; Keren, 1993). If participants who have already estimated the size of the appliances are then asked to estimate the energy consumption of the appliances, their energy estimates might be primed by their size estimates (e.g., Strack, Martin, & Schwarz, 1988). A strong correlation in this case would just reflect priming rather than any intuitive substitution of one attribute for the other. Similarly, participants who are asked to estimate both sets of attributes might realise the purpose of the study and respond differently than if they were unaware. If separate groups of participants estimate size and energy consumption, any influences come only from the knowledge, experience, and expectations that they bring into the study from their everyday lives (Greenwald, 1976; Keren, 1993). It is also slightly more realistic to only be judging one attribute or the other, rather than both variables in

succession (Kahneman & Frederick, 2002).

In the study reported in this paper, we tested for the use of the size heuristic using the heuristic elicitation design, which uses a between-participants design for data collection but the same method of correlation as Baird and Brier (1981, Experiments 2 & 3). This enabled us to investigate whether the correlation between size and energy consumption estimates was different when performed on within-participants data than on between-participants data.

1.3.2 Methods of correlation in heuristic testing. The method of correlation used to analyse participants' responses has been shown by Monin and Oppenheimer (2005, Study 3) to influence the strength of the correlation coefficient (the effect size of the study) (see also Nickerson, 1995). Averaging participants' data is a common way to make raw responses more manageable for analysis for both within-participants and between-participants data. Correlating average estimates is the only way to analyse between-participants data in which there are no intrinsic pairings between the two variables being correlated. In within-participants data, however, each participant provides estimates of both size and energy consumption for each appliance so there are more ways in which the data can be correlated.

When Baird and Brier (1981, Experiment 3) analysed their data, they first calculated the average estimate of size for each appliance and the average estimate of energy consumption for each appliance. They then correlated the average estimate of size and the average estimate of energy for each appliance (appliances were the unit of analysis rather than the participants). Monin and Oppenheimer (2005), however, found that averaging participants' ratings before correlating them led to higher correlation coefficients than if the raw ratings for each item were correlated per participant and then the coefficients averaged. They analysed the same within-participants dataset twice: Once by averaging the ratings before correlating them ($r = .52$) and once by correlating the ratings per participant then averaging the coefficients (mean $r = .25$) (Monin & Oppenheimer, 2005, Study 3). The difference between the coefficients was not

significant ($z = 1.71$, $p = .087^1$) and both were significantly larger than zero so the overall conclusion of the study did not change. The large differences in effect sizes as a result of using different methods of correlation, however, has concerning implications for how they might be interpreted by other researchers reviewing the energy heuristics literature or performing meta-analyses of the results in future.

Baird and Brier (1981) used the same method of correlation as in the heuristic elicitation design method: They averaged the estimates of size and of energy consumption for each appliance before correlating them. Schuitema and Steg (2005) took a different approach to correlation. They treated each participant-and-appliance combination as an individual case. With 16 appliances and 60 participants, their analysis contained 1560 cases ($16 \times 60 = 1560$). They then correlated all the size estimates with all the energy estimates. Each participant's size estimate for an appliance was correlated with that same participant's energy estimate for the same appliance but, instead of obtaining a separate coefficient for each participant and then averaging the coefficients, Schuitema and Steg (2005) obtained a single coefficient in one calculation. The correlation coefficient they obtained ($r = .67$), whilst reasonably large in itself was significantly smaller than the coefficients obtained by Baird and Brier (1981, Exp. 2: $r_s = .91$, $n = 19$, $z = 2.85$, $p = 0.004$, Exp. 3: $r_s = .93$, $n = 18$, $z = 3.27$, $p = 0.001$). There are several possible explanations for this difference in coefficient size. For example, they used different procedures to present the questions to the participants; Schuitema and Steg used a parametric correlation test instead of the non-parametric test that Baird and Brier (1981) used; Baird and Brier's findings could have been outliers and Schuitema and Steg's a regression to the mean. It is possible, though, that the difference in coefficient was down to the method of correlation used. In the study reported in this paper, we analysed the within-participants dataset using three different methods of correlation. This enabled us to test whether the different methods of

¹We used an online tool providing Fisher's (r -to- z transformation) (Lowry, 2015) to calculate the significance of differences between pairs of coefficients. Spearman's Rho correlations are Pearson's correlations using ranked data so it was appropriate to use the same test for both r and r_s (Myers & Sirois, 2004).

analysis led to different correlation coefficients and different conclusions about the strength of the relationship between people’s perceptions of the size and energy consumption of household appliances.

1.4 Study Overview

It is important to understand better how householders perceive their energy consumption if effective policies are to be designed to help them reduce household energy consumption. People use heuristics to help them make judgements and decisions in their everyday life. Previous evidence has shown that householders might use the size heuristic to help them judge the amount of energy consumed by their appliances but the studies varied in their findings and in their methods used.

The study in this paper tested for the use of the size heuristic using two different methods of data collection and three different methods of correlation to investigate whether the method affects the results. Table 1 summarises the four different study designs tested in this paper. The analysis methods are labelled according to the initials of the authors of the studies from which the method was taken. For simplicity, this paper refers to the method of correlation used by both Baird and Brier (1981) and the heuristic elicitation design (Kahneman & Frederick, 2002) as the BB analysis, regardless of whether it is used to analyse within-participants or between-participants datasets.

Design 1 tested whether a positive correlation between size and energy consumption would be obtained when data were collected using a between-participants design—the heuristic elicitation design (Kahneman & Frederick, 2002).

Design 2 tested whether there was a difference in the size of the correlation coefficient for data collected using a within-participants design, like Baird and Brier (1981, Experiments 2 & 3), compared with the data collected using a between-participants design in Design 1.

Design 3 tested whether the correlation method used by Schuitema and Steg (2005) would produce a positive correlation in our within-participants dataset and whether the coefficient was different from the coefficient obtained in Design 2.

Table 1

Summary of the four study designs tested in this paper.

Design	Data Collection	Data Analysis	Source of Method
1	between-participants	Correlate averaged estimates across participants (BB analysis) ^a	Kahneman and Frederick (2002)
2	within-participants	Correlate averaged estimates across participants (BB analysis)	Baird and Brier (1981, Experiments 2 & 3)
3	within-participants	Correlate raw estimates across participants (SS analysis)	Schuitema and Steg (2005)
4	within-participants	Correlate raw estimates per participant (MO analysis)	Monin and Oppenheimer (2005, Study 3)

^aThe heuristic elicitation method (Kahneman & Frederick, 2002) uses the same method of correlation as Baird and Brier (1981, Experiments 2 & 3) used but on between-participants data.

Design 4 tested whether the method used by Monin and Oppenheimer (2005, Study 3) would produce a positive correlation in our within-participants dataset and whether the coefficient was different from the coefficient obtained in Design 2.

A final analysis evaluated how closely (or not) participants' perceptions of the energy consumption of the appliances matched the actual average energy consumption of the appliances in UK households. Both the between-participants and within-participants datasets were analysed and both the BB analysis (used in Designs 1 and 2) and MO analysis (used in Design 4) were used to correlate the estimates with the actual consumption data.

The data for the within-participants and between-participants datasets were collected at the same time and using almost identical procedures so the Method for all the designs is described in the next section. The analyses, results, and conclusions are then reported separately for each design. The study reported in this paper was conducted in line with the university ethics regulations (University of Surrey, 2016).

2 Method

Two datasets were collected for the analyses reported in this paper: A between-participants dataset and a within-participants dataset. The within-participants dataset was analysed using three different methods of correlation. All measures, conditions, and data exclusions for this study are reported in this article.

2.1 Participants

Participants were recruited online using social media and a university recruitment system. Undergraduate Psychology students received a course token for participation; no other incentive was offered. Each participant was randomly allocated by the web-based survey software to either the within-participants group, the between-participants energy group, the between-participants size group, or one of three other groups. The data collected from the three other groups measured the use of other heuristics and will be analysed separately and reported elsewhere; they were not relevant to the research questions of this study, which focused specifically on comparing methods of testing for the size heuristic.

One participant in the between-participants size group was not included in the analyses because they did not complete the tasks. Two other participants were removed after initial analyses (one each from the between-participants size group and the within-participants group) when it became clear that they had not engaged in the tasks and had missed out (or estimated as zero) an unusually large number (up to 13) of appliances compared with other participants. Removal of these two participants did not significantly change any of the coefficients or conclusions so they are not mentioned further in this article².

In the between-participants energy and size groups together, 53 participants (27 in the energy group and 26 in the size group; 64% female) were recruited through social

²Intraclass correlations before removal for the within-participants dataset were .28 for energy consumption and .59 for size; for the between-participants dataset they were .22 for energy consumption and .58 for size. See the Results section of Design 4 for comparison.

media (70%, mainly Twitter and Facebook) and a university recruitment system (28%). Fifty-three percent were employed or self-employed; 33% were students; the remainder were retired, long-term sick or disabled, or looking after the home or family.

In the within-participants group, 30 participants (53% female) were recruited through social media (63%, mainly Twitter and Facebook) and a university recruitment system (23%). Sixty-seven percent were employed or self-employed; 27% were students; the remainder were retired, looking for work, or long-term sick or disabled.

2.2 Study Power

In Designs 1, 2, and 4, the household appliances were to be the unit of analysis (the cases) so the relevant sample size was the number of appliances rather than the number of participants. Power calculations (Mayr, Buchner, Erdfelder, & Faul, 2007) based on effect sizes from the literature (Baird & Brier, 1981; Schuitema & Steg, 2005) showed that a sample size of 8–25 would be required for an effect size (r_s) of .9–.6.

2.3 List of Appliances

Although power calculations specified that only 8–25 appliances would be required, it was important to include enough appliances to represent a range of energy consumption quantities and sizes and ensure variation in participants' responses. An initial pilot study included 44 appliances but participants complained that was too many on mobile phone screens. A subset of 30 appliances (see Table A1) was selected, giving a sample size of $n = 30$ in Designs 1, 2, and 4. In Design 3, the initial sample size was $n = 900$ because cases were each combination of 30 within-participants participants and 30 appliances (though this was reduced to $n = 895$ during analysis by listwise deletion of missing estimates). The appliances were taken from lists of energy-consuming appliances mentioned by interviewees in previous interviews (reported elsewhere).

2.4 Procedure

Participants were invited to participate in a study about the psychology of quantity to avoid them wondering why they were being asked to estimate the size of

household appliances. The procedures for participants in the between-participants and within-participants groups were almost identical. In the within-participants group, participants estimated, in counterbalanced order, both the size and energy consumption of the appliances, whereas in the between-participants group, participants estimated either the size or energy consumption. Although participants could skip appliances, they were prompted to respond to any that they had missed before continuing. Participants in the between-participants energy group took longer to complete the study than participants in the size group ($M_{\text{energy}} = 7.78$ minutes, $SD_{\text{energy}} = 11.60$; $M_{\text{size}} = 4.81$ minutes, $SD_{\text{size}} = 1.81$). Participants in the within-participants group took a mean of 7.80 minutes ($SD = 3.13$) except for one participant who left the study open in her browser for 4.5 days before submitting her responses.

Participants were asked to “Use the sliders to indicate the size of these objects, where 100 is the largest object and 0 is the smallest.” and to “Use the sliders to indicate how much energy you think each object uses in the average UK household in 1 year, where 100 is the object that uses the most and 0 is the least.”. Participants indicated their estimates by moving a slider along a scale with the minimum value of 0 and the maximum value of 100 (starting position = 0). Participants were requested to “Answer as quickly as you can and just say what you think. Don’t worry about being right or wrong.” to discourage them from doing online research or explicitly calculating the size or energy consumption.

Identical scales (0—100) were provided for both size and energy consumption estimates because participants were not expected to know the absolute values in kWh of the individual appliances (Attari et al., 2010; Kidd & Williams, 2008) and to avoid participants trying to explicitly calculate the size or energy consumption. Providing an upper bound to the scale meant it was not necessary to provide a reference point (modulus) against which all participants should make comparisons (Kahneman & Frederick, 2002) and which could have artificially anchored participants’ estimates (Attari et al., 2010; Frederick, Meyer, & Mochon, 2011; Schley & DeKay, 2015). Frederick et al. (2011) showed that choice of modulus probably varies between people

when estimating energy consumption and that artificially constraining the modulus can influence the size of the correlation. We expected participants to make their estimates relative to other appliances so the appliances were displayed in a random order for each participant, leaving participants free to choose their own reference points whilst being constrained to responding on the 0—100 scale.

As in the study by Schuitema and Steg (2005) (see also Schley & DeKay, 2015), participants were asked to judge annual energy consumption instead of hourly energy consumption as requested by Baird and Brier (1981, Experiments 2 & 3). In previous interviews, several participants emphasised the importance of considering seasonal variation and standby usage of appliances when estimating energy consumption. Annual estimates were requested to smooth out perceptions of varying usage of the appliances and to avoid uncertainty around responding for, say, summer versus winter usage.

After completing the estimation tasks, but before some demographic questions, all participants were asked an additional question to test for another heuristic that was unrelated to the research questions addressed in this article and so will be reported elsewhere. Its presentation after the estimation tasks ensured that it would have no impact on the results reported in this article.

2.5 Actual Energy Consumption and Size Data

The actual energy consumption data were calculated from the most recently available official UK energy statistics. The annual tonnes of oil equivalent per appliance were calculated from Table 3.10 and Table 3.12 of the Domestic dataset (DECC, 2015). Reliable and internally consistent annual energy consumption data could be obtained for only 16 of the 30 household appliances that participants estimated. The actual size (volume; cm^3) of each appliance was obtained from the online catalogue of a popular UK white goods website (Currys.co.uk).

The actual energy consumption and actual size data were standardised so that the highest consuming appliance consumed 100 units of energy and the largest appliance was 100 units in size (see Table A1). This made the data more manageable during the

analysis and more easily comparable with the estimates provided by participants on the 0–100 scale.

3 Results

For clarity, the analyses, results, and conclusions are reported separately for each design.

3.1 Design 1

3.1.1 Introduction. In Design 1, the data were collected and analysed according to the heuristic elicitation design method (Kahneman & Frederick, 2002; Kahneman & Tversky, 1996). The between-participants dataset was analysed using the BB analysis of averaging the size estimates and the energy consumption estimates per appliance and then correlating the averaged size and energy estimates (see Table 1). Based on the literature (Baird & Brier, 1981; Schuitema & Steg, 2005), it was hypothesised that size and energy estimates would be positively correlated (one-tailed).

3.1.2 Analyses. To prepare the between-participants dataset for analysis, the data were transposed so that the cases (rows) were the appliances and the variables (columns) were the responses to each question by each participant. The median size estimates and median energy consumption estimates were correlated for each appliance using a Spearman’s Rho test. The non-parametric test was most appropriate because a Kolmogorov-Smirnov test showed that the size estimates were not normally distributed ($D_{\text{energy}}[30] = .05$, $p = .200$; $D_{\text{size}}[30] = .19$, $p = .009$) and so violated parametric assumptions. Median estimates were used instead of means for comparability with Baird and Brier (1981).

3.1.3 Results. As hypothesised, there was a strong, positive correlation between the median estimates of size and energy consumption ($r_s = .78$, p [one-tailed] $< .001$; see Table 2). Although lower, the correlation coefficient was not significantly different from the coefficients Baird and Brier (1981) reported (Exp. 2, $r_s = .91$, $z = 1.53$, $p = .063$, Exp. 3, $r_s = .93$, $z = 1.9$, $p = .029$).

Table 2

Summary of correlation coefficients between size and energy consumption from each study design and, for comparison, from the literature.

Study	Method	Coefficient	n
Design 1	b-p; averaged estimates across participants (BB analysis)	$r_s = .78$	30
Design 2	w-p; averaged estimates across participants (BB analysis)	$r_s = .81$	30
Design 3	w-p; raw estimates across participants (SS analysis)	$r_s = .53$	895 ^a
Design 4	w-p; raw estimates per participant (MO analysis)	$r_s = .60$	30
Baird and Brier (1981, Exp. 2)	w-p; averaged estimates across participants (BB analysis)	$r_s = .91$	19
Baird and Brier (1981, Exp. 3)	w-p; averaged estimates across participants (BB analysis)	$r_s = .93$	18
Schuitema and Steg (2005)	w-p; raw estimates across participants (SS analysis)	$r = .67$	1560

Note. b-p = between-participants; w-p = within-participants. ^aAfter listwise deletion during the correlation.

3.1.4 Conclusion. According to the aim of the heuristic elicitation design that we used (Kahneman & Frederick, 2002), the strong, positive correlation between the estimates of the two groups of participants suggests that the participants who estimated the energy consumption of the appliances were actually estimating the size of the appliances (or vice versa).

3.2 Design 2

3.2.1 Introduction. In Design 2, the data were collected and analysed according to the method used by Baird and Brier (1981). The within-participants dataset was analysed using the BB analysis as in Design 1 (see Table 1). That is, the

average size estimates for each appliance were correlated with the average energy consumption estimates for each appliance. By keeping the data analysis method consistent between Designs 1 and 2, we could determine whether the data collection method (between-participants vs within-participants) could cause a difference in the correlation coefficient obtained. Based on the results of Baird and Brier (1981) and Schuitema and Steg (2005), it was hypothesised that the size and energy estimates would be positively correlated (one-tailed). It was also hypothesised that if the within-participants design influenced participants' estimates, the correlation between size and energy consumption estimates would be significantly different from in Design 1 (two-tailed).

3.2.2 Analyses. The within-participants dataset was prepared for analysis in the same way as the between-participants dataset in Design 1. Again, a Kolmogorov-Smirnov test showed that the size estimates were not normally distributed ($D_{\text{energy}}[30] = .10, p = .200$; $D_{\text{size}}[30] = .21, p = .002$) so Spearman's Rho was used for all correlations. Baird and Brier (1981) also used the Spearman's Rho test so using the same test enabled us to compare our correlation coefficient (r_s) with theirs.

3.2.3 Results. As hypothesised, there was a strong, positive correlation between the median estimates of size and energy consumption ($r_s = .81, p$ [one-tailed] $< .001$; see Table 2). This supported the results of Baird and Brier (1981) and Schuitema and Steg (2005), reflecting a real correlation—not just a consequence of their study design. There was no significant difference between the coefficients found using Design 2 and Design 1 ($z = 0.30, p = .764$).

The order in which participants in the within-participants dataset estimated the size and energy consumption of the appliances was counterbalanced. This meant that we could check for order effects using the same method as Strack et al. (1988), by correlating each counterbalance group separately and then comparing the two correlation coefficients. Although participants who responded to the size question first produced a slightly higher correlation coefficient ($r_s = .85, p$ [one-tailed] $< .001$) than participants who responded to the energy question first ($r_s = .81, p$ [one-tailed] $< .001$),

there was no significant difference ($z = 0.47$, $p = .638$) between the two. Both coefficients were very similar to the overall correlation coefficient for the within-participants dataset (for the median estimates per appliance in each dataset, see Table A1).

3.2.4 Conclusion. The type of data collection (within-participants vs between-participants) did not make a significant difference to the correlation coefficient obtained. It seems very unlikely that Baird and Brier (1981) and Schuitema and Steg (2005) obtained their strong positive correlations between size and energy consumption estimates as a result of using within-participants designs to collect their data.

3.3 Design 3

3.3.1 Introduction. In Design 3, the within-participants dataset was re-analysed using Schuitema and Steg (2005)’s method of correlation (SS analysis) in which a single set of raw size estimates was correlated with a single set of energy consumption estimates across all participants to obtain a single correlation coefficient (see Table 1). Based on the literature (Baird & Brier, 1981; Schuitema & Steg, 2005), it was hypothesised that size and energy estimates would be positively correlated (one-tailed). Based on the observed significant difference between the coefficients obtained by Baird and Brier (1981) and Schuitema and Steg (2005), it was hypothesised that the size of the coefficient would be smaller than that obtained in Design 2, which used the BB analysis (one-tailed).

3.3.2 Analyses. To prepare the within-participants dataset for analysis, the data were arranged with just two variables, energy consumption estimates and size estimates, so that each case was just two estimates from one participant for one appliance. In total, there were 900 cases (30 participants X 30 appliances), though through listwise deletion $n = 895$ in the correlation. When the dataset was prepared in this way, a Kolmogorov-Smirnov test showed that neither the energy consumption estimates nor the size estimates were normally distributed ($D_{\text{energy}}[898] = .10$, $p < .001$; $D_{\text{size}}[897] = .14$, $p < .001$) so Spearman’s Rho was used for all correlations, unlike in

Schuitema and Steg (2005)’s original study in which they appeared to use Pearson’s correlation coefficient.

3.3.3 Results. As hypothesised, there was a positive correlation between size and energy consumption estimates ($r_s = .53$, p (one-tailed) $< .001$; see Table 2). The coefficient was, however, significantly smaller than that obtained using Design 2 ($z = 2.75$, p [one-tailed] $= .003$), a similar pattern to the difference in correlation coefficients obtained by Schuitema and Steg (2005) compared with Baird and Brier (1981, Experiments 2 & 3).

As a final check, although a Kolmogorov-Smirnov test showed that our data were non-normal, we ran a Pearson’s correlation and obtained a very similar coefficient as when using the Spearman’s test ($r = .50$, p [one-tailed] $< .001$). It seems unlikely that the significant difference in effect sizes between Baird and Brier (1981, Experiments 2 & 3)’s and Schuitema and Steg (2005)’s studies was due to using different statistical tests.

3.3.4 Conclusion. This analysis showed that it was possible to obtain significantly different correlation coefficients by using different methods of correlation to analyse the same dataset for heuristics. It also seems unlikely that the difference between the coefficients obtained by Baird and Brier (1981, Experiments 2 & 3) and Schuitema and Steg (2005) was a consequence of regression to the mean.

3.4 Design 4

3.4.1 Introduction. In Design 4, the within-participants dataset was re-analysed again, this time using the method of correlation used by Monin and Oppenheimer (2005) (MO analysis; see Table 1). Based on their findings, it was hypothesised that the size and energy consumption estimates would still be positively correlated (one-tailed). It was also hypothesised that the correlation coefficient from Design 4 (using MO analysis) would be different from the coefficient from Design 2 (using BB analysis). It was not clear in which direction the difference would be (two-tailed).

3.4.2 Analyses. The within-participants dataset was re-analysed by correlating the size and energy consumption estimates of the appliances for each

participant separately (see Monin & Oppenheimer, 2005, Study 3). As in Design 2, the appliances were treated as cases and the participants and their estimates as variables. In Design 2, however, the energy and size estimates for each appliance were averaged before correlation. In Design 4, each participant's size and energy estimates of the appliances were correlated for the participant, then the 30 correlation coefficients (one per participant) were averaged to find the mean coefficient.

3.4.3 Results. As hypothesised (see Monin & Oppenheimer, 2005, Study 3), there was still a moderately good, positive correlation between size and energy consumption estimates. The coefficient was smaller this time but still significant (mean $r_s = .60$, 95% C.I. = [.52; .68]). The smaller effect size was probably due to a lack of consistency between participants in their estimates (Monin & Oppenheimer, 2005). We tested the intraclass correlation coefficient using a two-way random model for absolute agreement and found that participants were indeed inconsistent with each other in their estimates of energy consumption (.28), whilst being in moderately high agreement with each other in their estimates of size (.66). (Very similar intraclass correlation coefficients were found for the between-participants data of .22 for energy consumption estimates and .66 for size estimates.) The individual correlation coefficients of each participant also varied greatly (see Table B1). As Monin and Oppenheimer (2005, Study 3) found, the coefficients obtained by correlating averaged data (Design 2) versus raw data (Design 4) appeared to be very different though the difference was not statistically significant ($z = 1.59$, $p = .112$; see Table 2).

3.4.4 Conclusion. Although the difference between the correlation coefficients in Design 4 and Design 2 was not significant this time, using two different methods of correlation to analyse the data for heuristics produced quite different coefficients. This shows that researchers should not only be careful in choosing their method of correlation but also in clearly reporting it so that other researchers are not misled.

3.5 Comparing Perceptions With Average Energy Consumption

3.5.1 Introduction. We investigated how closely people’s perceptions of size and energy consumption matched the actual size and energy consumption of the appliances in average UK households, and the strength of the relationship between the actual size and energy consumption of appliances. This would give us more insight into participants’ use of the size heuristic when judging the energy consumption of appliances.

3.5.2 Analyses. We correlated the actual size and actual energy consumption data with each other and with the estimated energy consumption and estimated size data from both the within-participants and between-participants datasets. To further investigate the impact of using different methods of correlation to analyse the data, the correlations were performed using both the BB analysis (participants’ median estimates correlated with actual data) and MO analysis (participants’ raw estimates correlated, per participant, with actual data). Table 3 lists the coefficients for each dataset using each method of correlation. Appliances were treated as cases but only 16 appliances were included in correlations with the actual energy consumption data because it was available for only 16 of the 30 appliances.

3.5.3 Results. Figures 1 and 2 show the actual size and energy data with the estimates for the between-participants and within-participants datasets respectively. As Attari et al. (2010) and Baird and Brier (1981) found, participants seemed, on average, to underestimate energy consumption for the higher-consuming appliances but this could also reflect underlying differences between participants in their energy consumption estimates. Actual size and actual energy consumption correlated strongly with each other ($r_s = .90$, p [one-tailed] $< .001$).

Participants’ estimates of size were also strongly correlated with the actual size of the appliances; the MO analysis produced mean coefficients that were slightly smaller than the BB analysis produced but the difference was not significant (see Table 3. Between-participants dataset: $r_{sBB \text{ analysis}} = .91$, p [one-tailed] $< .001$; mean $r_{sMO \text{ analysis}} = .81$, 95% C.I. = [.75; .87]; $z = 1.41$, $p = .159$. Within-participants dataset: $r_{sBB \text{ analysis}}$

Figure 1. Actual energy consumption and size of appliances with between-participants average estimates.

$= .94$, p [one-tailed] $< .001$; mean r_{sMO} analysis $= .87$, 95% C.I. $= [.85; .89]$; $z = 1.49$, $p = .136$). Participants were, on average, accurate when estimating the size of appliances.

Participants' estimates of size strongly correlated with the actual energy consumption; again, the MO analysis produced mean coefficients that were slightly smaller than the BB analysis produced but the difference was not significant (see Table 3. Between-participants dataset: r_{sBB} analysis $= .90$, p [one-tailed] $< .001$; mean r_{sMO} analysis $= .80$, 95% C.I. $= [.74; .86]$; $z = 1.08$, $p = .280$. Within-participants dataset: r_{sBB} analysis $= .85$, p [one-tailed] $< .001$; mean r_{sMO} analysis $= .84$, 95% C.I. $= [.82; .86]$; $z = .10$, $p = .920$). Participants' perceptions of size were, on average, reliable to use as a heuristic when judging energy consumption.

Participants' estimates of energy consumption were strongly correlated with the actual energy consumption of the appliance when the BB method was used to correlate the median estimates of each appliance with the actual energy consumption of the appliance. When the MO analysis was used to correlate individual raw estimates of energy consumption with actual energy consumption, however, the positive correlations

Figure 2. Actual energy consumption and size of appliances with within-participants average estimates.

were lower, significantly lower in the case of the between-participants data (see Table 3. Between-participants dataset: $r_{sBB \text{ analysis}} = .85$, p [one-tailed] $< .001$; $r_{sMO \text{ analysis}} = .46$, C.I. = [.34; .58]; $z = 2.20$, $p = .028$. Within-participants dataset: $r_{sBB \text{ analysis}} = .87$, p [one-tailed] $< .001$; $r_{sMO \text{ analysis}} = .63$, C.I. = [.55; .71]; $z = 1.75$, $p = .080$). On average, participants' estimates of energy were fairly accurate, though not as accurate as their estimates of size. The lower correlations produced by the MO analysis, however, suggest that the accuracy of energy saving estimates varied between participants.

When the MO analysis was conducted on the within-participants dataset, both the correlation coefficients between energy consumption estimates and actual energy consumption and the correlation coefficients between the same people's estimates of size and energy consumption appeared to vary greatly between individuals (Table 3, column 7, and Table B1). An exploratory correlation of the two sets of coefficients for the same participants showed a strong, positive correlation of $r_s = .82$. This is probably unsurprising as both sets of coefficients were based on the participants' energy consumption estimates. It does confirm, however, that participants whose estimates of

Table 3

Actual and estimated size and energy consumption correlations.

	Between-participants				Within-participants			
	n	BB (r_s)	n	MO (r_s)	n	BB (r_s)	n	MO (r_s)
Estimated Size & Actual Size	30	.91	26	.81	30	.94	30	.87
Estimated Size & Actual Energy	16	.90	26	.80	16	.85	30	.84
Estimated Energy & Actual Energy	16	.85	27	.46	16	.87	30	.63

Note. BB = BB analysis; MO = MO analysis; correlations that involved actual energy consumption data were conducted on only the subset of 16 appliances for which the data were available.

size and energy consumption correlated highly also tended to be more accurate in their estimates of energy consumption.

3.5.4 Conclusion. The strong, positive correlations produced by the BB analysis method between average size estimates and the actual energy consumption of appliances suggest that a size heuristic would be useful in judgements of energy consumption, especially when considered alongside the strong, positive correlations between actual size and actual energy consumption, and size estimates and actual size. The correlations between average energy consumption estimates and the actual energy consumption of appliances (i.e., the accuracy of participants' energy consumption estimates) are slightly lower than would be expected based on the accuracy of the size estimates if people were relying purely on size to judge energy consumption. This suggests that, as Baird and Brier (1981) concluded, while size might play a part in people's judgements of energy consumption, it is not the only cue that people use.

The correlations using the MO analysis method suggest a slightly different interpretation. Only some participants were very accurate in their energy consumption estimates; other participants varied greatly in their accuracy (see Table 3, column 7). We also found that participants whose size and energy consumption estimates correlated strongly also tended to be more accurate in their energy consumption estimates. One possible interpretation of this finding is that participants who used the

size heuristic were more accurate in their estimates of energy consumption.

4 Discussion

It is essential to understand how householders perceive the energy consumption in their homes if behaviour change interventions and policies are to be effective. Simply providing householders with more and more information does not appear to be helping them to reduce their energy consumption. Identifying the heuristics used in energy consumption judgements would help to develop a better understanding of householders' mental models. That understanding can then be used to design more effective policies and, in future studies, to assess whether householders' perceptions of energy consumption and saving are associated with actual energy-saving behaviour. Kahneman and Frederick (2002) defined a simple methodology, the heuristic elicitation design, that researchers can use to clearly identify whether potential heuristics are used by householders. Although previous studies have already found evidence that people use the size heuristic when judging the energy consumption of household appliances, their results and methodologies varied. The study reported in this paper tested whether people use the size heuristic and whether the effect would persist under different study designs. We used two different methods of data collection (between-participants and within-participants) and three different methods of correlation to analyse the data (correlating averaged estimates across participants, correlating raw estimates across participants, and correlating raw estimates per participant).

In Designs 1 and 2, we found significant, strong, positive correlations of average size and energy consumption estimates regardless of whether the data were collected using a between-participants design or a within-participants design. This confirmed previous findings by Baird and Brier (1981, Experiments 2 & 3) and Schuitema and Steg (2005) that perceptions of size are strongly related to perceptions of energy consumption. It also confirms that the within-participant study design did not influence the correlations by priming or other means and, according to the attribute substitution model (Kahneman & Frederick, 2002), participants were substituting size when judging

energy consumption.

In Designs 3 and 4, we found that re-analysing the within-participants dataset with other methods of correlation produced significant, moderately strong, positive, but smaller coefficients. This suggests that the significant difference between the Baird and Brier (1981, Experiments 2 & 3) and Schuitema and Steg (2005) coefficients could have been due to the different methods of correlation that they used. The results of Design 4 also showed the relevance of individual differences in participants to understanding their perceptions of energy consumption. Whilst fairly consistent in their estimates of size, participants were much less consistent in their estimates of energy consumption, possibly reflecting the relative difficulty of the task of judging the energy consumption of an appliance compared with estimating its physical size.

Finally, we found that the size and energy consumption of appliances are strongly correlated in reality so the size heuristic would help householders to accurately judge the energy consumption of their appliances. On average, participants were less accurate in their energy consumption estimates than might have been expected from the accuracy of their size estimates. When analysed individually, however, there appeared to be large variation between individual participants in the accuracy of their energy consumption estimates. Participants who appeared to use the size heuristic (showed strong correlations between their size and energy consumption estimates) tended to be more accurate in their energy consumption estimates. Taken together, this suggests that householders vary in their mental models of the energy consumption of their appliances. Some participants had mental models that matched the actual energy consumption of appliances, while others did not.

4.1 Implications for Studying Householders' Perceptions of Energy Consumption

We confirmed the findings of Baird and Brier (1981) and Schuitema and Steg (2005) that there is at least a moderately strong positive correlation between people's estimates of the size and energy consumption of household appliances. Our finding

supports the suggestion, from both quantitative and qualitative studies, that some (but not necessarily all) householders use the size heuristic when judging the energy consumption of their appliances (Baird & Brier, 1981; Chisik, 2011; Schuitema & Steg, 2005; van den Broek, 2016).

4.1.1 The impact of methodology on results. We showed that different methods of correlation can produce different sizes of coefficients—the researcher’s choice of correlation method influences the effect size of the study. We hope that more research is conducted into how householders perceive and make sense of energy consumption and saving, including investigating the use of potential heuristics in their judgements. It is therefore important that researchers clearly report which correlation method they have used so that other researchers can interpret and compare their findings accurately.

The choice of method for collecting data appeared to be less important to the results of our study. We obtained the same strong, positive correlation for both the between-participants dataset and the within-participants dataset. We also found no order effects for the within-participants dataset. It is not clear why but it is possible that there were so many (randomly-ordered) appliances to estimate that participants did not remember their estimates of the first attribute for each appliance when estimating the second attribute and so were not primed or otherwise influenced by the study design (Li, Johnson, & Zaval, 2011, also found no order effects using a within-participants design to test attribute substitution). Another possible explanation is that size is one of a few special attributes that are always cognitively accessible to people and cannot be primed (Kahneman & Frederick, 2002). People automatically perceive the size of objects around them because they use size to help make perceptual judgements such as the distance of objects from them. Participants would not be able to help but substitute size when asked to judge the more difficult attribute of energy consumption. Research on other attributes potentially used as heuristics in energy consumption judgements, using both methods of data collection, would help to verify this.

It is possible that estimating energy consumption over one year instead of the much shorter time frame of one hour that was used by Baird and Brier (1981,

Experiments 2 & 3) caused different perceptions but it did not seem to affect our results. Baird and Brier (1981) and Schuitema and Steg (2005) used different time frames but, using a single time frame, we found a similar difference in results between the BB analysis and SS analysis methods as was produced by Baird and Brier (1981) and Schuitema and Steg (2005).

4.1.2 Causality in heuristic testing. The study reported in this paper, like those of Baird and Brier (1981) and Schuitema and Steg (2005), was correlational and so could not statistically confirm a causal relationship between participants' perceptions of size and energy consumption when judging household energy consumption. Baird and Brier (1981, Experiments 2 & 3) and Schuitema and Steg (2005) concluded a causal relationship based on specific errors made by their participants who appeared to be misled by appliances that consumed more or less energy than their size might suggest. Size is also more concrete, visible, and familiar than energy so it seemed plausible that it could influence perceptions of energy consumption rather than the other way round.

Of course, it is possible that the relationship between size estimates and energy consumption estimates was mediated by some other factor. For example, Gabe-Thomas, Walker, Verplanken, and Shaddick (2016), using cluster analyses, showed that participants appeared to think about appliances according to their location (in particular, the kitchen) rather than by their physical size. Most of the largest appliances in a typical household are generally found in the kitchen so people might use the location of the appliance rather than its physical size as a heuristic when judging its energy consumption. This explanation does not account for the errors reported by Baird and Brier (1981) and Schuitema and Steg (2005) though.

Another explanation could be that as the size and energy consumption of appliances are strongly related in reality and as our participants were all accurate in their estimates of size, if they were also accurate in their estimates of energy consumption even without using the size heuristic, they would produce strong, positive correlations between their size and energy consumption estimates. This demonstrates a limitation of the heuristic elicitation design in testing for the use of some energy

consumption heuristics. The method is useful in identifying potential heuristics but more experimental approaches (beyond the scope of this paper) would be necessary to verify causal relationships.

4.2 Size and Other Heuristics in Energy Judgements

It seems likely that size is relevant to people when judging energy consumption because size is easy to observe and cannot be avoided in everyday visual perception (Kahneman & Frederick, 2002). Other characteristics of appliances are likely to also be relevant to people’s judgements and should be tested in future studies. Size was tested in this study because it is the most reported energy judgement heuristic in the literature—probably because size perceptions are relatively easy to measure.

4.3 Individual Differences in Perceptions of Energy Consumption

An important finding of this study is that individual differences should be considered and measured in future studies that investigate people’s perceptions of household energy consumption. We found large variations between individual participants in their use of the size heuristic and the accuracy of their energy consumption estimates. Cognitive motivation and cognitive ability might be relevant to their perceptions. For example, people with higher numeracy skills tend to be more accurate in their judgements of energy consumption (Attari et al., 2010; Schley & DeKay, 2015). Kempton (1986) and Revell and Stanton (2014) also found variation between people in their mental models of central heating systems.

5 Conclusions

As Baird and Brier (1981) and Schuitema and Steg (2005) found, there is a moderately strong, positive correlation between people’s perceptions of the size of household appliances and their perceptions of the energy consumption of those appliances. The method of correlation chosen by researchers to analyse the relationship between participants’ estimates can influence the effect size of the study. It is important, therefore, that researchers choose carefully which method to use and also

report clearly their study design so that other researchers can interpret and compare studies. The method of data collection did not, in itself, make a difference to the strength of the correlation. Using a within-participants design with a per-participant method of correlation, however, revealed that individual differences are important to whether people might use the size heuristic.

While the original, between-participants heuristic elicitation design method should be used to quickly and cleanly identify likely heuristics, the MO analysis, with its sensitivity to individual differences, is valuable in providing insight into how people perceive energy consumption. It is clear from our findings that people vary in how they perceive the energy consumption of the appliances in their homes. It is clear from the literature that providing householders with increasing amounts of information is not helping them to reduce their energy consumption and save energy. Understanding how people perceive energy consumption is essential to being able to design effective energy-saving interventions that actually work.

6 References

- Attari, S., DeKay, M. L., Davidson, C. I., & Bruine de Bruin, W. (2010). Public perceptions of energy consumption and savings. *Proceedings of the National Academy of Science*, *107*, 16054–16059. doi: 10.1073/pnas.1001509107
- Baird, J. C., & Brier, J. M. (1981). Perceptual awareness of energy requirements of familiar objects. *Journal of Applied Psychology*, *66*, 90–96. doi: 10.1037/0021-9010.66.1.90
- Burgess, J., Harrison, C. M., & Filius, P. (1998). Environmental communication and the cultural politics of environmental citizenship. *Environment and Planning*, *30*, 1445–1460.
- Burgess, J., & Nye, M. (2008). Re-materialising energy use through transparent monitoring systems. *Energy Policy*, *36*, 4454–4459. doi: 10.1016/j.enpol.2008.09.039
- Chisik, Y. (2011). An image of electricity: Towards an understanding of how people perceive electricity. In *Interact 2011* (pp. 100–117).
- DECC. (2015). *Energy consumption in the UK (ECUK): Domestic data tables [Data file]*. Retrieved from <https://www.gov.uk/government/organisations/department-of-energy-climate-change/series/energy-consumption-in-the-uk>
- DECC. (2016). *Energy consumption in the UK* (Tech. Rep.). Department for Business, Energy & Industrial Strategy. Retrieved from www.gov.uk/beis
- Delmas, M. A., Fischlein, M., & Asensio, O. I. (2013). Information strategies and energy conservation behavior: A meta-analysis of experimental studies from 1975 to 2012. *Energy Policy*, *61*, 729–739. doi: 10.1016/j.enpol.2013.05.109
- European Union. (2017a). *Climate strategies & targets*. Retrieved from https://ec.europa.eu/clima/policies/strategies_en
- European Union. (2017b). *Final energy consumption by sector - Eurostat [dataset]*. Retrieved from <http://ec.europa.eu/eurostat/web/products-datasets/-/tsdpc320>

- European Union. (2017c). *Paris Agreement / Climate Action*. Retrieved from https://ec.europa.eu/clima/policies/international/negotiations/paris_en
- Frederick, S., Meyer, A. B., & Mochon, D. (2011). Characterizing perceptions of energy consumption. *Proceedings of the National Academy of Sciences of the United States of America*, *108*, E23. doi: 10.1073/pnas.1014806108
- Gabe-Thomas, E., Walker, I., Verplanken, B., & Shaddick, G. (2016). Householders' mental models of domestic energy consumption: Using a sort-and-cluster method to identify shared concepts of appliance similarity. *Plos One*, *11*, 1–15. doi: 10.1371/journal.pone.0158949
- Greenwald, A. G. (1976). Within-subjects designs: To use or not to use? *Psychological Bulletin*, *83*, 314–320.
- Kahneman, D., & Frederick, S. (2002). Representativeness revisited: Attribute substitution in intuitive judgment. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics and biases: The psychology of intuitive judgement* (pp. 49–81). Cambridge, England: Cambridge University Press.
- Kahneman, D., Schkade, D., & Sunstein, C. R. (1998). Shared outrage and erratic awards: The psychology of punitive damages. *Journal of Risk and Uncertainty*, *16*, 49–86. doi: 10.1016/0167-6687(98)90012-9
- Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review*, *80*, 237–251.
- Kahneman, D., & Tversky, A. (1982). On the study of statistical intuitions. *Cognition*, *11*, 123–41.
- Kahneman, D., & Tversky, A. (1996). On the reality of cognitive illusions. *Psychological review*, *103*, 582–91.
- Kempton, W. (1986). Two theories of home heat control. *Cognitive Science*, *10*, 75–90.
- Kempton, W., & Montgomery, L. (1982). Folk quantification of energy. *Energy*, *7*, 817–827. doi: 10.1016/0360-5442(82)90030-5
- Keren, G. (1993). Between- or within-subjects design: A methodological dilemma. In G. Keren & C. Lewis (Eds.), *A handbook for data analysis in the behavioral*

- sciences: Methodological issues* (pp. 257–272). Hillsdale, NJ: Erlbaum.
- Kidd, A., & Williams, P. (2008). *The Talybont Trial, exploring the psychology of smart meters* (Tech. Rep.). The Prospectory. Retrieved from http://www.fozienergy.com/Talybont_Trial_UK_Efergy_2008.pdf
- Kollmuss, A., & Agyeman, J. (2002). Mind the gap: Why do people act environmentally and what are the barriers to pro-environmental behavior? *Environmental Education Research*, 8, 1469–5871. doi: 10.1080/1350462022014540
- Li, Y., Johnson, E. J., & Zaval, L. (2011). Local warming: daily temperature change influences belief in global warming. *Psychological Science*, 22, 454–459. doi: 10.1177/0956797611400913
- Lowry, R. (2015). *Significance of the difference between two correlation coefficients*. Retrieved from <http://vassarstats.net/rdiff.html>
- Mayr, S., Buchner, A., Erdfelder, E., & Faul, F. (2007). A short tutorial of G*Power. *Tutorials in Quantitative Methods for Psychology*, 3, 51–59. doi: 10.1037/0096-1523.32.4.932
- Monin, B. (2003). The warm glow heuristic: When liking leads to familiarity. *Journal of personality and social psychology*, 85, 1035–1048. doi: 10.1037/0022-3514.85.6.1035
- Monin, B., & Oppenheimer, D. M. (2005). Correlated averages vs. averaged correlations: Demonstrating the warm glow heuristic beyond aggregation. *Social Cognition*, 23, 257–278. doi: 10.1521/soco.2005.23.3.257
- Myers, L., & Sirois, M. J. (2004). Spearman correlation coefficients, differences between. In *Encyclopedia of statistical sciences* (chap. 12). Hoboken, NJ, USA: John Wiley & Sons, Inc. doi: 10.1002/0471667196.ess5050.pub2
- Nickerson, C. A. E. (1995). Does willingness to pay reflect the purchase of moral satisfaction? A reconsideration of Kahneman and Knetsch. *Journal of Environmental Economics and Management*, 28, 126–133.
- Norman, D. A. (1983). Some observations on mental models. In D. Gentner & A. L. Stevens (Eds.), *Mental models* (pp. 7–14). Hillsdale, NJ: Lawrence Erlbaum

Associates.

- Revell, K. M. A., & Stanton, N. A. (2014). Case studies of mental models in home heat control: Searching for feedback, valve, timer and switch theories. *Applied Ergonomics*, 45, 363–378. doi: 10.1016/j.apergo.2013.05.001
- Schley, D. R., & DeKay, M. L. (2015). Cognitive accessibility in judgments of household energy consumption. *Journal of Environmental Psychology*, 43, 30–41. doi: 10.1016/j.jenvp.2015.05.004
- Schuitema, G., & Steg, L. (2005). Percepties van energiegebruik van huishoudelijke apparaten. In A. E. Bronner, P. Dekker, E. d. Leeuw, K. d. Ruyter, A. Smidts, & J. E. Wieringa (Eds.), *Ontwikkelingen in het marktonderzoek. jaarboek markt onderzoek associatie2005* (pp. 165–180). Haarlem: De Vrieseborch.
- Strack, F., Martin, L. L., & Schwarz, N. (1988). Priming and communication: Social determinants of information use in judgments of life satisfaction. *European Journal of Social Psychology*, 18, 429–442. doi: 10.1002/ejsp.2420180505
- Tversky, A., & Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review*, 90, 293–315.
- University of Surrey. (2016). *University of Surrey research ethics*. Retrieved from <http://www.surrey.ac.uk/research/our-approach/ethics>
- US Energy Information Administration. (2013). *Heating and cooling no longer majority of US home energy use*. Retrieved from [https://www.eia.gov/todayinenergy/detail.php?id=10271&src=%E2%80%B9ConsumptionResidentialEnergyConsumptionSurvey\(RECS\)-f3](https://www.eia.gov/todayinenergy/detail.php?id=10271&src=%E2%80%B9ConsumptionResidentialEnergyConsumptionSurvey(RECS)-f3)
- US Energy Information Administration. (2015). *Passenger travel accounts for most of world transportation energy use*. Retrieved from <https://www.eia.gov/todayinenergy/detail.php?id=23832>
- US Energy Information Administration. (2016). *Use of energy in the United States*. Retrieved from https://www.eia.gov/energyexplained/index.cfm?page=us_energy_use
- van den Broek, K. L. (2016). *Antecedents of energy literacy and energy saving*

behaviour: A mixed methods approach (Unpublished doctoral dissertation)

(Unpublished doctoral dissertation). University of Bath, UK.

Appendix A

Annual UK energy consumption and size of appliances and median estimates

Table A1

Annual UK energy consumption and actual volume of appliances with participants' median estimates of energy and size.

Appliance	Actual Energy ^a	Actual Volume ^b	B-P Estimates		W-P Estimates	
			Energy	Size	Energy	Size
an energy-saving lightbulb	2.45	0.07	15.00	5.00	6.00	5.00
an old-style filament lightbulb	5.21	0.05	40.00	6.00	35.00	5.00
a halogen lightbulb	5.83	0.03	34.00	5.00	14.00	3.00
a games console	12.88	5.48	35.00	20.00	29.00	17.00
a laptop	14.72	0.53	36.00	20.50	27.00	18.50
a microwave	27.30	9.52	37.50	35.00	33.00	30.00
a Virgin/Sky/Tivo box	32.52	0.66	27.00	17.50	23.00	12.00
a 30-inch LCD TV	33.44	26.62	44.00	41.00	35.00	42.50
a kettle	44.17	2.04	39.00	11.00	39.00	16.00
a fridge	45.40	63.85	50.00	67.00	47.00	62.50
an oven	46.01	48.03	57.00	67.50	71.00	53.50
a washing machine	54.60	62.44	60.00	69.50	51.00	60.00
a chest freezer	69.33	100.00	45.00	78.50	44.00	65.50
an upright freezer	69.63	48.74	57.00	68.50	51.50	74.00
a dishwasher	82.52	71.69	50.00	69.00	58.00	58.00
a tumble dryer	100.00	74.64	65.00	69.50	67.00	59.50
a bathroom extractor fan		0.02	23.00	13.50	17.50	10.00
a central heating boiler		22.31	64.00	58.50	75.00	52.00
a DAB (digital) radio		0.68	16.00	14.00	14.00	11.50
a free-standing electric heater		27.72	68.50	30.00	61.50	30.00
a hairdryer		1.67	30.00	10.50	32.00	11.50
a lawnmower		32.86	41.00	44.00	29.50	46.50
a mobile phone		0.01	20.00	6.00	18.00	5.00
a smoke alarm		0.20	7.50	7.00	3.00	5.00
a tablet/iPad		0.06	19.00	11.50	21.00	9.50
a vacuum cleaner		28.58	32.00	36.50	37.00	31.00
a wifi/internet router		0.62	22.00	11.00	15.00	7.00
an electric blanket		3.86	35.00	22.50	33.50	20.00
an electric shower		0.92	49.50	29.00	56.00	24.50
an iron		1.33	32.00	13.00	26.50	13.50

Note. B-P = between-participants dataset; W-P = within-participants dataset. ^aThe actual energy consumption of each appliance was standardised so that the highest consuming appliance consumed 100 units of energy. The *tonnes of oil equivalent* were multiplied by 3067.484662577. The number of units is rounded to 2 decimal places for clarity in this table. ^bThe actual size of each appliance was standardised so that the largest appliance was 100 units in size. The cm^3 volume was multiplied by 0.000238039.

Appendix B

Individual correlation coefficients for Design 4

Table B1

Correlation coefficient for each within-participant participant's estimate of perceived size and perceived energy consumption.

Participant ID	Correlation coefficient (r_s)
1	.39
2	.66
3	.69
4	.61
5	.70
6	.46
7	.82
8	.84
9	.11
10	.68
11	.59
12	.50
13	.62
14	.56
15	.80
16	.85
17	.81
18	.11
19	.58
20	.60
21	.23
22	.65
23	.67
24	.41
25	.67
26	.88
27	.39
28	.65
29	.80
30	.70

Appendix C

Individual correlation coefficients for the accuracy test with MO method

Table C1

Correlations between each participant's estimates and the actual size and energy consumption of the appliances.

Participant ID	Estimated and Actual Size		Estimated Size and Actual Energy		Estimated and Actual Energy	
	b-p	w-p	b-p	w-p	b-p	w-p
1	.82	.94	.88	.88	.16	.50
2	.88	.91	.87	.86	-.15	.82
3	.81	.95	.84	.84	.74	.74
4	.91	.84	.81	.80	.59	.57
5	.88	.90	.89	.86	.00	.68
6	.45	.89	.61	.86	.58	.56
7	.88	.84	.87	.82	.15	.71
8	.79	.79	.86	.79	.80	.92
9	.88	.86	.91	.80	.59	.45
10	.80	.84	.86	.82	.78	.62
11	.87	.95	.87	.91	.70	.66
12	.79	.88	.83	.88	.20	.42
13	.95	.80	.84	.80	.68	.58
14	.89	.90	.89	.90	.25	.71
15	.89	.88	.76	.84	.65	.83
16	.25	.85	-.01	.84	.74	.74
17	.84	.89	.83	.88	.80	.81
18	.91	.90	.90	.81	.49	-.11
19	.82	.76	.86	.82	.62	.68
20	.86	.89	.88	.78	.39	.61
21	.76	.81	.82	.83	.34	.33
22	.74	.92	.74	.84	.43	.76
23	.94	.88	.83	.88	.45	.58
24	.78	.81	.80	.84	.65	.34
25	.90	.96	.85	.89	.79	.82
26	.88	.70	.83	.78	.27	.78
27	—	.89	—	.87	-.29	.53
28	—	.88	—	.85	—	.59
29	—	.91	—	.94	—	.90
30	—	.89	—	.86	—	.84

Note. b-p = between-participants dataset; w-p = within-participants dataset.