Response-times and subjective complexity of food choices: A web-based experiment across 3 countries

Please do not quote without authors' permission

Forthcoming in

Social Science Computer Review

December 2021

Rossella Atzori[a]*

Andrea Pellegrini[b]

Ginevra Virginia Lombardi[c]

and

Riccardo Scarpa[d,e,f]

[a] Department of Economics and Business, University of Cagliari, Italy

[b] Centre for Business Intelligence and Data Analytics, Business School, University of Technology Sydney, Australia

[c] Department of Economics and Management, University of Florence, Italy

[d] Department of Business Administration, University of Verona, Italy

[e] Durham University Business School, Durham University, United Kingdom

[f] Waikato Management School, University of Waikato, New Zealand

*Corresponding author: rossella.atzori@unica.it - Department of Economics and Business, University of Cagliari, via Sant'Ignazio, 17, IT-09123, Cagliari, Italy

Declaration of Conflicting Interest

The Authors declare that there is no conflict of interest.

Acknowledgements

This work was supported by Regione Toscana (Italy). The second author was funded to visit the University of Technology Sydney by the Swiss National Science Foundation [P2TIP1_187933].

Response-times and subjective complexity of food choices: A web-based experiment across 3 countries

Abstract

Accurate collection of response times is one of the main advantages of web-administered stated choice experiments and it can be thought of as a behavioral indicator of cognitive effort. We use data from a food choice experiment administered across three countries and estimate a panel Mixed Multinomial Logit Model to obtain individual-specific utility weights. These are used to construct two utility-based measures of contextual choice complexity, which are combined with subjective measures of cognitive resources as well as indicators of opt-out selection. We first develop and then test hypothesized effects of complexity at the level of single choice task and choice sequence on response times. By using a log-linear random effects model with choice task response-time as dependent variable we isolate these effects from other background variables. Results suggest that as our measures of complexity increase so do response times and such effects are robust across the three countries. We argue that these results broadly support the validity of web-based choice surveys to measure food preference. We suggest that computers can help improve survey design by implementing algorithms to improve the overall efficiency of choice tasks design, for example by using adaptive design algorithms that control cognitive challenges in accordance with the respondent's predicted ability to tackle cognitive effort.

Keywords

Response time; Choice Experiments; Computer-based surveys; Cognitive effort; Food choices

1 Introduction

Stated preference (SP) data are widely used by academics and marketing practitioners as a valid source of information for eliciting consumer preferences for different goods and services (Ben-Akiva & Lerman, 1985; Luce, 1957; McFadden, 1974, 1980, 1999). SP data are collected with various interview modes: in-person face-to-face, mail-in questionnaires, telephone surveys, computer-assisted and web-based modes, (MacMillan et al., 2002) however, since the early 1990s, online surveys have rapidly gained popularity among researchers for their speed, lower cost and greater reliability of data capture and storage (Couper & Miller, 2008; Lindhjem & Navrud, 2011; Merten and Ruch, 1996; Scott et al., 2011). During the Covid-19 period, web-based surveys have become even more attractive due to social-distancing restrictions and the broader uptake of web-based interactions. This survey administration mode is appealing as it enables users to implement sophisticated questionnaires that can rely on enhanced reality and infographics, as well as respondent-friendly designs that are comparatively more informative and often engage respondents interactively (adaptive methods). There are also drawbacks such as the high non-response rate and often unrepresentative sample composition. This may be due to differences in internet access or use across socio-economic segments (e.g., across ages, education levels, types of job, social classes, etc.) and to panel composition and respondent self-selection. The latter is mostly based on self-recruitment mechanisms and is often unrepresentative of the target population (Hillygus et al., 2014). Despite such disadvantages, internet-based surveys remain attractive due to their many advantages and most practitioners currently agree that online survey methods as an SP data gathering tool are "here to stay". This study focusses on the analysis of response times (henceforth abbreviated to RT) and its relationships with the cognitive challenge respondents face in stated choice surveys. Specifically, RT belong to that class of ancillary data defined as client-side or server-side paradata, which are automatically collected during computer-based or web-based surveys and can be used to explore behavioral patterns that are either difficult or impossible to examine in other survey modes (Callegaro et al., 2014; Heerwegh, 2003; McClain et al., 2019; Safir et al., 2001). Here we will focus on RT as a proxy for cognitive effort in stated choice and on a selection of meaningful determinants of such effort, which are under the control of the survey developer.

Aside from RT (Fan and Yan, 2010; Matjašič et al., 2018) client-side paradata include keystrokes (Kreuter et al., 2010), mouse clicks, tracking of mouse hovering and pausing over specific areas on the display (Kaye-Blake et al. 2009) and response changes (Couper, 2000). Server paradata, instead, generally refers to webpage visits and time stamps. Prior applications based on client-side paradata were mainly focused on keystroke data collection to assess survey errors (Kreuter et al., 2010). Since then, paradata from web surveys have received growing attention as a promising source of information on questionnaire completion and enabled further exploration of individuals' decision-making processes. The goal is to enhance survey techniques and respondents' experience (Fan & Yan, 2010; Stern, 2008) and recent progress has been made in our understanding of the role of respondents' attitude towards the use of such data (Kunz & Gummer, 2019). Within the broader category of paradata, RT measurements have attracted the attention of researchers from a variety of disciplines, particularly for their ability to monitor respondent behaviors, such as degree of survey engagement and signaling of survey-satisficing (Conrad et al., 2017). RT has been referred to as the time each survey respondent takes to answer either a question or a group of questions, down to times relating to each single interaction with the computer interface, from keystrokes on the keyboard, to submission entry of responses to each question. Alone or in combination with other paradata, RT data is frequently used to test the quality and validity of responses (Höhne et al., 2017), to better classify response patterns and to analyze the affordability of survey questions (see Henninger & Plieninger, 2020). Researchers' interest in exploring the relationship between RT and individual cognitive effort within decision making processes predates web-surveys. Psychologists investigated RT (Luce, 1986) and time to think in face-to-face interviews (see Fazio, 1990; Fischhoff, 2005; Svedsater, 2007; Whittington et al., 1992). Lenzner et al. (2010), on the other hand, explored the determinants of RT to enhance survey design. For example, from the perspective of linguists who care about textual presentation, suggesting threshold approaches for selection based on RT distribution (Mayerl 2013). More broadly, RT affect interview duration, with well documented effects (Crawford et al. 2001; Galesic & Bosnjak 2009) on rates of participation, completion and self-selection. More recently, Gummer and Rossman (2015) used multi-level analysis to disentangle the separate explanatory roles of respondent characteristics, survey tool, and the interaction between respondent's attitudes and web-survey design on RT.

SP data are often collected by using discrete choice experiments (DCE) surveys, as we do in this study. One of the earliest DCE papers (Holmes et al. 1998) found that RT affects the strength of stated preferences. Cook et al. (2007, 2012) report that longer thinking times correlates with respondents expressing lower willingness to pay (hence forth WTP) and lower quantity demanded. When respondents given longer time to think (one night) provided WTP estimates that on average were 40% lower. Campbell et al. (2018) used data from an online DCE survey to study preferences for honey types, origin and methods of production. Using a random utility framework, they find evidence of a complex and equivocal relationship between 'fast' and 'slow' RT and estimates of marginal WTP for honey attributes. Such evidence chimes with their previous results on choice of recreational fishing site (Campbell et al., 2017), where RT was used to inform sample partitions of latent class models. Ambiguous links between RT and choice behavior was previously reported by

Haaijer et al. (2000), who used two different datasets and found confirmation of opposite theoretical predictions with regards to the time taken to select the specific alternative of "no-choice". Theoretical expectations on the length of this time differ; with Dhar (1997) expecting it to be longer than the time taken to choose real alternatives, and Johnson and Orme (1996) expecting it to be shorter.

RT collected in online DCE surveys have been meaningfully correlated with respondent characteristics (e.g., age, educational level, etc.), question complexity (Yan & Tourangeau, 2008) and multitasking activities (Höhne, 2020). Vista et al. (2009) find that RT correlates solely with age, whilst Brown et al. (2008) found significant correlation patterns with age as well as other demographic variables. Bonsall and Lythgoe (2009) emphasize the correlation of RT with age and education of respondents---an obvious resource for cognitive effort---and, in accordance with other studies, they find that RT increases as a function of cognitive effort requirements, as measured by the complexity of the choice tasks used in the experimental design (Höhne et al., 2017; Lenzner et al., 2010; Vörös & Rouet, 2016). This is also a main focus in our study.

As initially reported in Haaijer et al. (2000), DCE studies frequently report evidence of the impact of RT on choice complexity, respondent fatigue along the sequence of choice tasks, and on the magnitude of utility error variance (or equivalently, "scale heterogeneity"), as assumed in random utility models. This had been theoretically predicted in the mid-90's (Espinoza-Varas & Watson, 1994), and has been broadly and regularly corroborated since by many (e.g. Arentze et al., 2003; Bech et al., 2011; Caussade et al., 2005; DeShazo & Fermo, 2002; Rose & Black, 2006, Scarpa et al. 2011, amongst others). More recently, Börger (2016) also found that RT is linked to both higher Gumbel error scale (lower utility variance) and higher accuracy in WTP estimates.

Hess and Stathopoulos (2013) considered simultaneous estimation of RT measurement functions and random utility parameters by using hybrid choice model specifications. Unfortunately, empirical samples in this category of models result in extremely low statistical likelihoods and hence have limited value in inference (Caspar and Kroesen, 2014; Vij and Walker 2016). It is important to note that either an insensitivity of RT to the complexity of choice tasks or any significant pattern of theoretically implausible correlations would seriously undermine the behavioral validity of stated preference data. Because validity is a very sensitive and controversial topic in hypothetical choice data of this type (see for example Johnston et al. 2017), progress in this line of investigation is much called for and of interest to a wide spectrum of survey practitioners and academic disciplines. Contributing to such progress is the focus of our study.

Against the above background, we contribute with an econometric analysis of RT determinants using the same DCE web-survey instrument in three countries: the U.K., Germany and Italy. The international scope is functional to ascertaining the generalizability of online response behavior across countries with substantially different food histories, cultures, customs and attitudes. The specific food product object of choice is the classic hot-dog sausage (sometimes called 'Frankfurters' or 'Würstel sausages'). Our underlying hypothesis being that structural determinants of the cognitive effort required during DCE web-surveys impact on RT in the same manner across countries. Our analysis intends to isolate the effects on RT of individualized and contextualized measures of cognitive challenge at the choice task level. Specifically, we derive subjective utility differences and entropy measures from random utility models of the panel mixed logit type (Revelt & Train 1998) and use them as independent variables in a log-linear random effects model to explain observed RT. We aim at separating these effects from those of well-established variables that measure cognitive constraints and resources, such as respondent's socio-demographics, survey context and environmental factors. We focus on qualitative food choice because of the growing popularity of DCE in this area of research, but we expect our study to be informative to analysts engaged in DCE web surveys of all types.

The rest of the paper is organised as follows. The next section outlines the theory and the hypotheses that characterize this study, followed by the methodological section. Section 4 describes the data collection process of our empirical application on hot-dog sausages, whilst Section 5 reports the results and their interpretation. Section 6 concludes the paper and suggests some future avenues for investigation.

2 Theory and Hypotheses

In DCEs respondents are asked to identify their preferred alternatives from a mutually exclusive set of qualitatively different alternatives arranged according to an experimental design. Higher levels of complexity in these choice tasks induce respondents to increase their cognitive effort if a rational decision coherent with their preferences is to be made. This paper builds on the assumption that cognitive effort for rational choice requires time[1]. As such, RT can be seen as its external manifestation and can be empirically and precisely measured during computer-based surveys.

---

[1]An increasing body of research allows for deviations from the rational choice theory, and that individuals may adopt a variety of alternative heuristics that simplify the decision process, for example by ignoring some attributes (Scarpa et al., 2009) or alternatives (Campbell & Erdem, 2019), revealing lexicographic preferences (Hess et al., 2010; Tversky, 1969) or comparing options on the basis of given thresholds (Swait, 2001; Cantillo & Ortúzar, 2006). In all these cases individuals will ignore some information of the decision problem in order to make a quicker and more efficient choice. In such circumstances, a quicker decision may still be rational. However, this study follows the standard consumer choice theory, assuming that individuals will accurately compare the bundles of product in all their characteristics. This being the case, a similar decision-making process necessitates time.

2.1 Choice complexity, response time and cognitive effort

There are several factors that can be hypothesized to affect cognitive effort. We divide these factors in subjective and choice-contextual. Previous results (Duquette, 2010) showed that it is the interaction of such factors that can explain self-rated levels of cognitive effort in choice task execution. The subjective factors are further distinguished in socio-economic characteristics of the subject (e.g., age, gender, income etc.) and in cognitive constraints (e.g., education level, familiarity with the food choice task, type of web-interface, etc.). The choice-contextual are factors linked to the structure of the choice task at hand, which can be controlled by researchers. Several such constructs have been proposed in the choice modelling literature in the early 2000's (DeShazo & Fermo, 2002; Swait & Adamowicz, 2001a, 2001b), taking the form of various measures of diversity and complexity of the choice task facing the respondent. To compute these measures (entropy, utility differences, standard deviations of utilities, etc.) requires evaluation of fitted values of utilities for alternatives in the choice tasks and of forecast of selection probabilities. These, in their turn, depend on estimates of a vector of preference weights ($\boldsymbol{\beta}$), which need to be estimated from the observed choice data. In our case, these measures are individualized by using the fitted utility and probability values for each individual respondent, as derived conditionally on the panel of observed choices. The estimation results for the MMNL model and the details of the model specification (draws, distributional assumptions, etc.) are provided in the Appendix available online.

We denote respondents by the subscript n and RT for the choice task t by τ_{nt}, while RT for the total sequence of choices faced by a respondent by τ_n. To measure choice complexity at the choice task level we follow Swait and Adamowicz (2001a, 2001b) and use the normalised Shannon Index of fitted probability diversity of each choice task t:

(1) $\widehat{H}_{nt} = \dfrac{\sum_j \hat{p}_{jnt} \ln(\hat{p}_{jnt})}{\ln(k)}$

where k is the number of alternatives in each choice task (in our case 3). However, unlike others before us, we compute $\hat{p}_{jnt}$ using the fitted logit probability of alternative selection (see equation 9 below) derived from the individual-specific $\widehat{\boldsymbol{\beta}}_n$ (see equation 14 below).

A higher value of $\widehat{H}_{nt}$ is observed for choice tasks with similar choice probabilities for the three alternatives. The value is highest when probabilities are identical:

(2) $\hat{P}_{nt}(j = 1) = \hat{P}_{nt}(j = 2) = \hat{P}_{nt}(j = 3) = 1/3.$

Provided the expected benefits derived from selecting the truly preferred alternative are sufficiently high to compensate the cost of cognitive effort (see Cameron & DeShazo, 2010), the respondent is expected to deploy more cognitive effort, spending a longer time compared to when selection probabilities are markedly different.

As a second measure of choice complexity, we use the absolute value of the difference between utilities of food purchase alternatives:

(3) $\widehat{\Delta}V_{nt} = |\hat{V}_{jnt} - \hat{V}_{int}|$,

which are also "subjectivized" to the specific respondent by using $\widehat{\boldsymbol{\beta}}_n$. When subjectively fitted utility differences across purchase alternatives are large, the preferred alternative becomes more readily apparent to the respondent, thereby reducing cognitive effort and τ_{nt}. This measure is important to capture cognitive effort when respondents are considering selecting one of the two purchase alternatives in the task.

We also measure choice complexity for the entire sequence of $T = 10$ choice tasks. To do so we use the average values of both measures over the sequence:

(4a) $\bar{H}_n = \frac{1}{T}\sum_t \widehat{H}_{nt}$, and

(4b) $\overline{\Delta V_n} = \frac{1}{T}\sum_t \widehat{\Delta}V_{nt}$.

The last contextual factor of cognitive burden in choice emerges from an interaction between the minimum thresholds of the acceptability criteria by the respondent and the purchase alternatives. Given the presence of a "no-buy" option in all choice tasks, it is expected that when both of the experimentally designed food purchase alternatives fail to satisfy the subjective thresholds the no-buy option would be selected. This may happen, for example, when both food alternatives are too expensive for the budget of the respondent. In this case the respondent does not need to commit the same level of cognitive effort to choose between the two food product profiles, and trade-off all attributes. A similar effect on RT (τ_{nt}) might also be observed when respondents decide to opt-out as a form of protest; perhaps due to a lack of interest in the valued good or maybe because they are unwilling to pay for the proposed alternative. It is also worth noting that online questionnaires are often distributed through survey research firms to paid panelists. Many of these have an incentive to complete many surveys in a short time. Often these respondents may pay insufficient attention to the attributes due to the speed with which they go through the sequence of choice tasks.

As such, τ_{nt} values are expected to be on average smaller if the opt-out alternative is selected for these reasons. To measure this effect, another choice complexity indicator is constructed as a dummy variable at the choice task level as:

(5) $opt.out_{nt} = 1(y_{nt} = no.buy)$

The measure for the entire sequence is simply the sum of these dummy variables across the panel of choices:

(6) $opt.out_n = \sum_t 1(y_{nt} = no.buy)$

The higher the number of “opt-out” choices the lower the cognitive effort required along the sequence and the shorter the time taken to formulate a valid stated choice response.

2.2 Behavioral hypotheses on contextual factors

Adequate cognitive engagement is necessary to generate a valid response at each choice task of the DCE sequence. Everything else equal, a complex choice task requires a higher cognitive effort and hence, likely, a longer RT. Our hypotheses are therefore implemented with regards to the following derivatives on the subjective measure of complexity:

Null Hypothesis 1:

$$\frac{\partial \tau_{nt}(H_t)}{\partial H_t} > 0 \text{ versus the alternative } \frac{\partial \tau_{nt}(H_t)}{\partial H_t} \leq 0, \text{ and}$$

Null Hypothesis 2:

$$\frac{\partial \tau_{nt}(\Delta V_t)}{\partial \Delta V_t} < 0 \text{ versus the alternative } \frac{\partial \tau_{nt}(\Delta V_t)}{\partial \Delta V_t} \geq 0, \text{ and}$$

Null Hypothesis 3:

$$\frac{\partial \tau_{nt}(opt.out_t)}{\partial opt.out_t} < 0 \text{ versus the alternative } \frac{\partial \tau_{nt}(opt.out_t)}{\partial opt.out_t} \geq 0.$$

Similar hypotheses are formulated for the effects on RT for the entire choice task sequence, τ_n (the estimation results of the log-linear model for τ_n are included in the Web Appendix).

2.3 Behavioral hypotheses on subjective factors

Generally speaking, those relating to cognitive resource constraints, such as lower level of education, lower familiarity with food purchase decisions, survey taken at a later time of the day, (although this might also depend on the circadian type of the respondent, such as her chronotype, degree of morningness/eveningness (see Blatter & Cajochen, 2007) which we did not gather data for), etc. can be formulated as RT being longer whenever the cognitive resource is scarcer (i.e., lower educational attainment, familiarity and later in the day). These will be discussed in more detail in the results section.

3 The empirical framework

In this section, we illustrate the methodological framework employed to evaluate the determinants of variation of individual RTs in our web-based food survey. In doing so, we first present the details of the Mixed Multinomial Logit (MMNL) model estimated on the DCE data, from which we derived the subjective preference weights used to compute the contextual determinants of cognitive efforts discussed in the previous section. In our DCE the three mutually exclusive alternatives of each choice task comprise two unlabelled food choices plus a no-buy option. The MMNL model is used to evaluate the individual-specific estimates of the utility parameters, which are then used to compute two contextual measure of complexity for each choice task ($\widehat{H}_{nt}$, $\widehat{\Delta}V_{nt}$) and respondent sequence ($\widehat{H}_n$, $\widehat{\Delta}V_n$). These are at the core of our research hypotheses 1 and 2. Such measures are subsequently included, along with other subjective determinants of cognitive effort, in the analysis of the variability of RTs. We empirically evaluate our hypotheses by implementing two semi log-linear models, in which the dependent variables are τ_n and τ_{nt} . The semi-log specification is justified by the positive nature of the dependent variables.

3.1 The MMNL model

The MMNL is often described as the 'workhorse' of discrete choice analysis under preference heterogeneity, i.e. under unobserved random taste variation across decision-makers. The MMNL can approximate with any level of accuracy any preference structure consistent with random utility maximization theory (see McFadden & Train, 2000). In this model, utility of alternative j for web survey respondent n is assumed to be linear and additive in the attribute vector $\mathbf{x}_j$ used to describe the hotdog sausage (see Figure 1) and weighted by the respondent's utility weights in the vector $\boldsymbol{\beta}_n$. At the level of the individual respondent n, in choice task t, the conditional (observable) utility of a hotdog sausage alternative j is assumed to take the form:

(7) $V_{jnt}(\boldsymbol{\beta}_n) = \boldsymbol{\beta}'_n \mathbf{x}_{jnt}$.

While the decision-maker knows her total utility from each hotdog sausage alternative, the researcher can only approximate it on the basis of the observables $\mathbf{x}_{jnt}$. So, to compute overall utility for alternative j researchers assume that the unobservable component ϵ_{jnt} is random. Such a component is assumed to be independent of both the attribute vector describing the food alternative $\mathbf{x}_{jnt}$ and the vector of utility weights of the respondent $\boldsymbol{\beta}_n$:

(8) $U_{jnt} = V_{jnt}(\boldsymbol{\beta}_n) + \epsilon_{jnt} = \boldsymbol{\beta}'_n \mathbf{x}_{jnt} + \epsilon_{jnt}$.

With a distributional assumption on the unobservable component ϵ_{jn} – typically that of being distributed i.i.d. Gumbel – and given the vector of tastes of the individual $\boldsymbol{\beta}'_n$ the selection probability for alternative j at choice task t in the sequence is logit:

(9) $\Pr(i|\boldsymbol{\beta}_n) = \dfrac{\exp(\boldsymbol{\beta}'_n \mathbf{x}_{int})}{\sum_j \exp(\boldsymbol{\beta}'_n \mathbf{x}_{jnt})}$.

To obtain more accurate estimates of individual preferences (the values in $\boldsymbol{\beta}_n$) within the DCE survey each web-respondent n is asked a sequence of T choices, yielding a vector $\mathbf{y}_n^T$ of observed stated choices. Assuming independence across choices in the sequence up to the same preference vector $\boldsymbol{\beta}_n$, the joint probability of the sequence of choices is:

$$(10)\ \Pr(\mathbf{y}_n^T|\boldsymbol{\beta}_n) = \prod_t^T \Pr\left(i_t|\boldsymbol{\beta}_n\right) = \prod_t^T \frac{\exp\left(\boldsymbol{\beta}'_n\mathbf{x}_{int}\right)}{\sum_j \exp\left(\boldsymbol{\beta}'_n\mathbf{x}_{jnt}\right)}.$$

Obviously, preferences $\boldsymbol{\beta}_n$ vary across people. So, some distribution law needs to be invoked to account for such variation in the population from which the sample is derived. This is parametrically defined as $f(\boldsymbol{\beta}_n|\boldsymbol{\theta})$, where $\boldsymbol{\theta}$ is a vector of unknown parameters (e.g., means and variance-covariance matrix) that can be estimated with observations on $\mathbf{y}_n^T$ and $\mathbf{x}_n^T$. That is, with sequences of T choices made by each respondent n in a sample of size N the joint sample likelihood becomes:

$$(11)\ \Pr(\mathbf{y}_n^T|\boldsymbol{\beta}_n) = \int_{\boldsymbol{\beta}_n} \left[\prod_t^T \frac{\exp\left(\boldsymbol{\beta}'_{nt}\mathbf{x}_{int}\right)}{\sum_j \exp\left(\boldsymbol{\beta}'_{nt}\mathbf{x}_{jnt}\right)}\right]^{y_{jt}} f(\boldsymbol{\beta}_n|\boldsymbol{\theta})\, d\boldsymbol{\beta}_n.$$

The above is the well-known representation of the MMNL for repeated choices (Revelt & Train, 1998). The integral does not have a closed form, so in estimation it is approximated by simulation by averaging over a large number R of probabilities computed at (pseudo-)random draws of $\boldsymbol{\beta}_n^r$:

$$(12)\ \widetilde{\Pr}(y_n^T|\boldsymbol{\beta}_n) = \frac{1}{R}\sum_{r=1}^R \prod_t^T \left[\frac{\exp\left(\boldsymbol{\beta}'_n\mathbf{x}_{jnt}\right)}{\sum_i \exp(\boldsymbol{\beta}'_n\mathbf{x}_{int})}\right]^{y_{jt}}.$$

Once population estimates $\widehat{\boldsymbol{\theta}}$ have been obtained, the analyst can derive individual-specific estimates for the means of each respondent's distributions of $\boldsymbol{\beta}_n$ conditional on the observed pattern of $\mathbf{y}_n^T$ choices (see chapter 11 in Train, 2009 for details). Every individual making an identical set of choices and facing with the same set of choice tasks will have the same individual-specific distribution.

Using Bayes' rule this conditional distribution and its means are:

$$(13)\ h(\boldsymbol{\beta}|\boldsymbol{y}_n\boldsymbol{x}_n,\boldsymbol{\theta}) = \frac{\Pr(\boldsymbol{y}_n|\mathbf{x}_n,\boldsymbol{\beta}_n)\, f(\boldsymbol{\beta}|\boldsymbol{\theta})}{\Pr(\boldsymbol{y}_n|\mathbf{x}_n,\boldsymbol{\theta})} \xrightarrow{yields} \bar{\beta} = \int \boldsymbol{\beta} h(\boldsymbol{\beta}|\boldsymbol{y}_n\mathbf{x}_n,\boldsymbol{\theta})\, d\boldsymbol{\beta}.$$

Individual mean values for such conditional distributions can be approximated by simulation for each respondent as:

$$(14)\ \widehat{\boldsymbol{\beta}}_n = \sum_r w^r \boldsymbol{\beta}\text{, where } w^r = \frac{\widetilde{\Pr}(\boldsymbol{y}_n|\boldsymbol{\beta}_n)}{\sum_r \widetilde{\Pr}(\boldsymbol{y}_n|\boldsymbol{\beta}_n)},$$

and are considered more informative than population means to predict behavior at the individual respondent level, and crucially in our case, also individual evaluations of utility differences and perceived choice task complexity, in eq. (1) and (3). As pointed out by an anonymous reviewer, the former quantities can also be computed for each individual and choice task conditional on the observed sequence of individual choice task responses, as in eq. (14). For example, for eq. (1) it would be:

$$(14a)\ \widehat{\overline{H}_n} = \sum_r w^r \overline{H}_n\text{, where } w^r = \frac{\widetilde{\Pr}(\boldsymbol{y}_n|\boldsymbol{\beta}_n)}{\sum_r \widetilde{\Pr}(\boldsymbol{y}_n|\boldsymbol{\beta}_n)},$$

which is obviously different from eq. (14). Expectation values of non-linear functions of random variables----such as eq. (1) and (3)---differ from values of non-linear functions computed at the expected values of random variables. The implicit interpretation in our computation is as if respondents used their average individual preference to assess complexity, instead of using the average of their complexity evaluations over their individual distribution of preferences. We do not investigate this issue further here and leave this matter to future studies to explore.

3.2 Log-linear random effects models

To test our hypotheses and explain RT variation we use two semi log-linear specifications. In the first, the dependent variable is τ_n, the time spent by respondents to complete the entire sequence of T choice tasks. We report those estimates in the appendix. In the second, the dependent variable is τ_{nt}, the time respondents spent on each choice-task and the estimates are detailed here, given their richer interpretation.

Consider a respondent n that completes the DCE in τ_n minutes. The semi log-linear model for this observation can be written as follows:

$$(15)\ \ln(\tau_n) = \mu + \boldsymbol{\gamma}'\mathbf{z}_n + \varepsilon_n$$

In the above equation, μ is the intercept, $\mathbf{z}_n$ is a vector of explanatory variables associated with individual n, ε_n is an error term i.i.d. with $(0, \sigma_\varepsilon^2)$ and $\boldsymbol{\gamma}$ is a vector of parameters to be estimated. $\mathbf{z}_n$ can be further decomposed into $\mathbf{z}_n = [\mathbf{s}_n \mathbf{q}_n]$ where $\mathbf{s}_n$ is a sub-vector of socio-economic characteristics and choice-contextual factors of the respondent n and $\mathbf{q}_n$ is a sub-vector that accommodates the two summary measures of choice complexity along the sequence and reported in equation (4), which we hypothesize to be relate to cognitive effort deployed in the choice sequence, of which τ_n is the observable manifestation. To account for each respondent n subjective perception these are calculated at the individual means of the preference coefficients of the fitted MMNL model (equation 14). The reader will note that the left-hand side is expressed in the form of natural logarithm. Doing so ensures that we are able to control for potential asymmetry and skewness patterns naturally present in τ_n (see Figure 3).

The second model retains the specification shown in Equation (9), except for the fact that the dependent variable now represents the time observed to be taken by each respondent n to perform each choice task t, denoted as τ_{nt}. This gives rise to a semi log-linear panel model as follows:

(16) $\ln(\tau_{nt}) = \alpha + \boldsymbol{\varphi}'\mathbf{z}_{nt} + \omega_n + \eta_{nt}.$

As can be seen from Equation (16), the proposed modelling approach encompasses two stochastic components: the term η_{nt} is a normally distributed i.i.d error term with distributions $N(0, \sigma_\eta^2)$, whereas ω_n is a respondent-specific random effect i.i.d. $N(0, \sigma_\omega^2)$. The latter varies across respondents and remains constant over choices made by the same respondent. With this error structure, we can account for the source of variation in the RT that may arise at the choice task level as well as between individuals. All parameter estimates are obtained by Maximum Likelihood. We note, as suggested by an anonymous reviewer, that efficient estimations of the parameters in both choice and log-linear models could be obtained in a simultaneous maximization of the joint sample probabilities. We leave this rather complex extension to future studies to implement, noting that under the right model specification, our 2-stage approach is asymptotically consistent if not efficient, and represents a contribution in line with 2-stage approaches used in previous research.

4 Data

The data used in this study come from a DCE survey carried out between August and September 2018. The survey was designed using a cloud-based software platform (Qualtrics™) and was optimized for both computers and mobile devices[2]. The questionnaire was administered online to 2,862 respondents across three European countries: 1,200 in Italy, 820 in Germany and 842 in the United Kingdom. Data are provided by two highly reputable, but distinct survey research firms: the first collected all the observations for Italy and 400 of the observations from Germany; the second provided the remaining observations for Germany and all those from the U.K. Some observations

[2] One question type included in the survey (i.e. the matrix tables used for eliciting Likert scale questions) is not deemed ideal for guaranteeing an optimal survey's mobile experience. The use of this format was carefully assessed in the pre-test and it was not considered a source of concern for the overall quality of the survey. In any case, this potential shortcoming does not apply to the present work, as RTs used for analysis refer only to optimised questions.

were removed due to unrealistic values in one field (the respondent's age). The final dataset comprises 2,855 respondents: 1,198 in Italy, 817 in Germany and 840 in the U.K.

The objective of the survey was to elicit consumer preferences towards a common meat product, namely hot dog sausages. However, the attributes of policy interest relate to these being produced with innovative sustainable food processes (e.g. using natural preservatives, using extensive rearing systems, and being produced in mountain regions). The questionnaire gathered respondents' food purchase habits, their DCE responses, their individual attitudes towards food (e.g. tendency to innovate and attitude to health, traditions, trust) and towards specific dimensions of meat products (e.g. use of natural or synthetic preservatives and the respondents' perception of the risk associated with their use) and general socio-economic information.

The DCE used a *C*-efficient design (Scarpa & Rose, 2008), with 60 choice situations divided in 6 blocks of 10 choice tasks each (T = 10). Each choice task corresponds to a choice between two alternatives and the status quo. Figure 1 shows the description of the eight attributes characterizing the hot dog sausages, whereas Figure 2 presents an example of a choice card.

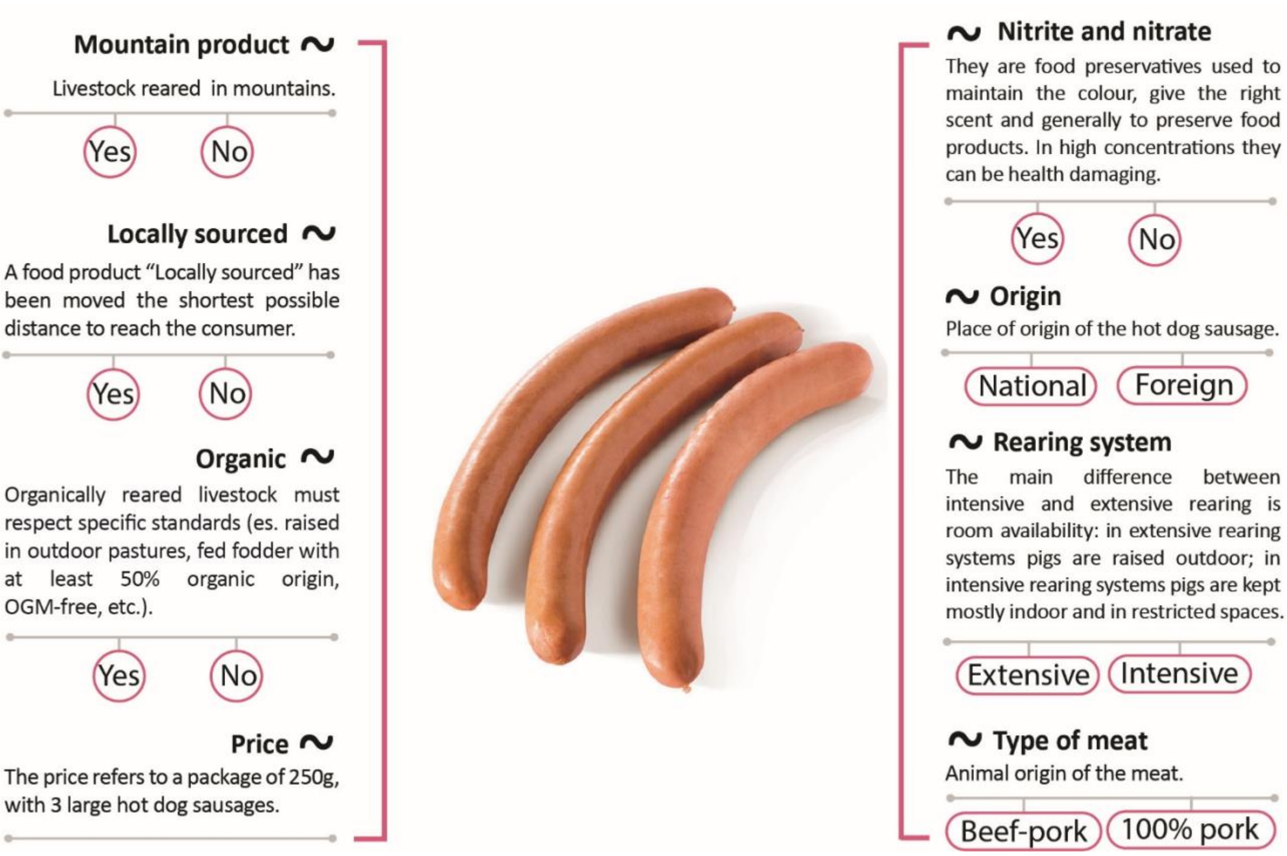

Figure 1 Attributes and levels of the choice experiment

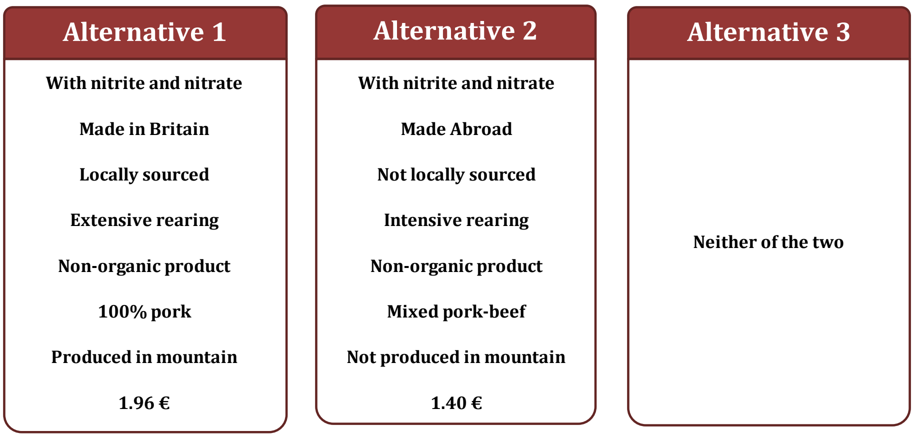

Figure 2 Example of choice task as showing on the computer screen

5 Results and discussion

5.1 Descriptive statistics

Table 1 depicts the descriptive statistics pooled across countries for τ_n and τ_{nt}.

Table 1 Summary statistics for RT variables (minutes)

	Variable	Mean	St.Dev.	Min.	Percentiles 2.5	25	50	75	97.5	Max
All obs.	τ_n	3.653	6.121	0.361	0.787	1.798	2.665	3.828	12.685	180.527
	τ_{nt}	0.365	1.796	0.000	0.044	0.122	0.209	0.353	1.232	177.202
95% obs.	τ_n	3.055	1.749	0.786	0.953	1.842	2.664	3.748	7.882	12.627
	τ_{nt}	0.271	0.207	0.044	0.058	0.126	0.209	0.342	0.870	1.228

The DCE RT to complete the 10 choice tasks (τ_n) ranged between 0.4 and 180 minutes, with a mean of approximately 3.7 minutes and a median of 2.67. On the other hand, the RT for each choice task (τ_{nt}) ranged between 0 and 177 minutes with a mean of approximately 22 seconds (2.55 in natural logs) and a median of 12.5 seconds.

A further aspect that appears from the differences between average and median values is that τ_n and τ_{nt} have highly right-skewed distributions, which is a common occurrence in RT from web surveys. Values in the upper tail of the distribution of τ_n (e.g., above 20 minutes to complete the DCE) suggest that some respondents completed the exercise (or the task) while engaging in other tasks. In contrast, values in the lower tail of the distribution are very close to zero and may signal that the respondent did not take the minimum necessary time required to read all the attribute levels and consider all necessary or expected trade-offs. The effect of outliers on normality is evident in Figure 3 where we report QQ plots of the natural log of RT.

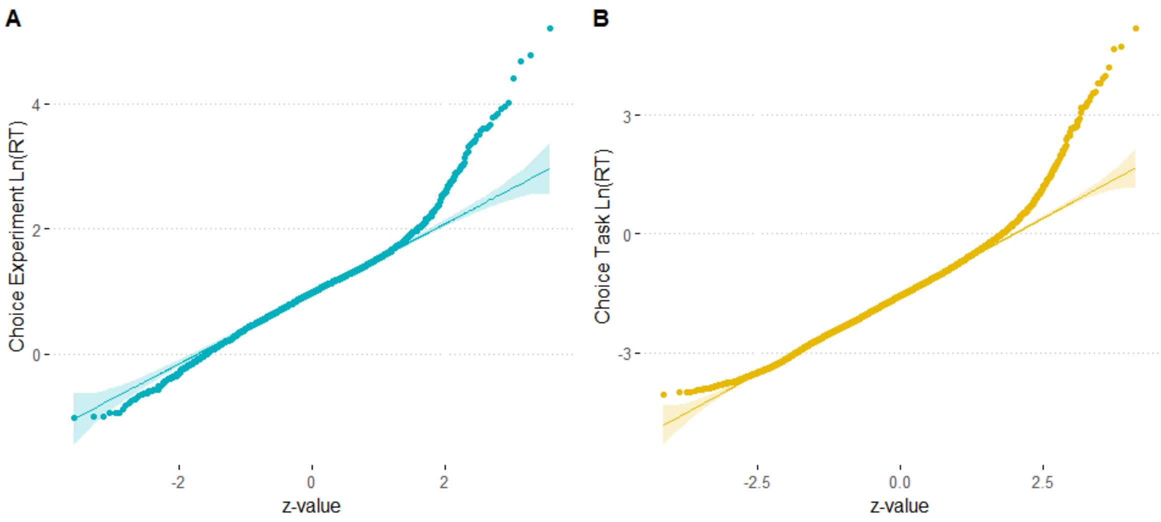

Figure 3 QQ plots of the natural log of RT

Keeping observations with very short and very long RT might produce biased estimates, and outliers are customarily removed from estimation datasets (Gummer & Rossman, 2015; Zhang & Conrad, 2014; Greszki et al. 2015). We removed observations outside the interval between the 2.5th and the 97.5th percentiles from our analyses. This produced an unbalanced panel in the semi-log regression explaining the variation of $\ln(\tau_{nt})$ the final dataset of which includes 27,121 choices, while that for $\ln(\tau_n)$ comprises 2,712 individuals. Once outliers are removed, the empirical sample distributions of the RT of each country (Figure 4) still show skewness but only minor differences across countries are apparent. The Italian sample has slightly higher mean value for both RTs and a fatter tail. The null hypothesis of country samples being drawn from a distribution with the same population mean is strongly rejected (p-value <0.001 Wilcoxon test) across all pairs, except for the UK-Germany for τ_{nt} (p-value 0.43). The pair-wise nonparametric test for the null of equality of continuous distributions is also always strongly rejected (p-value <0.001 Kolomogorov-Smirnoff) except for UK-Germany (p-value 0.11), but this time for τ_n. Finally, the nonparametric Kruskal-Wallis rank test for the three samples originating from the same distribution is always rejected.

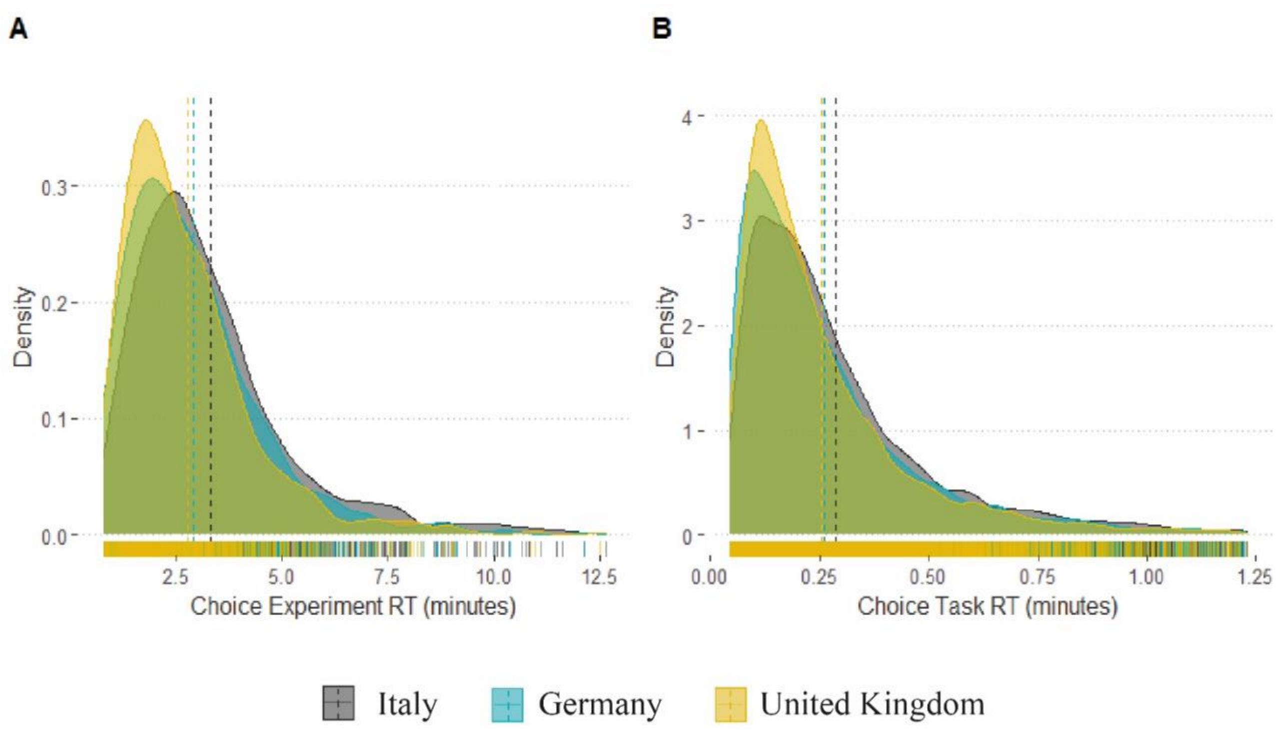

Figure 4 Distributions of RT by country

Evidence of different distributions does not negate that the underlying variables can have marginal effects of determinants with similar dimension and identical sign. To investigate these issues, and test our hypotheses, we first run regressions explaining the variation for the sequence of choices τ_n (see the Web Appendix for the results) and then those for the analogous model in panel form explaining the richer variation across all choice tasks τ_{nt}.

Sample statistics for the relevant socio-demographics used in the regressions are described in the Web Appendix (Table W.2), whereas the characteristics of choice tasks are discussed in more detail in the following section.

5.2 RT at the choice-task level

Table 2 shows the results of the semi-log (unbalanced) panel regression with random effects for RT at the single choice-task level (τ_{nt}) estimated on 27,121 observations. Altogether this much richer sample provides a more articulated set of results on RT, as opposed to the model applied to the entire DC sequence, the full estimates of which are reported in Web Appendix. Model (1) presents a specification including only socio-demographic and contextual variables $\mathbf{s}_n$, Model (2) includes Shannon's entropy measures $\widehat{H}_{nt}$ for each choice task and its interaction with a dummy for Italian respondents, and Model (3) includes the absolute difference between utilities $\widehat{\Delta V}_{nt}$ as well as a dummy variable denoting that the alternative selected was the no-purchase $opt.out_{nt}$. All our null hypotheses are strongly supported by the results of the panel data regressions.

Model (1)-(3) suggest that the relationship between age groups and τ_{nt} is positive and highly significant. Being of older age has the single strongest effect. Respondents aged 35-50 spend approximately 15-16% longer in completing a choice task than their younger counterparts (the benchmark is younger than 35). These effects are larger for those aged 50-65 (37-42%) and for those older than 65 (47-53%). Considering the average time of 16 seconds to answer a choice task, a 50% increase in the RT amounts to 8 additional seconds to complete the task. Older respondents may either need longer to process the necessary information or are less familiar with (or skilled at) navigating web surveys. On average, men are faster than women, with a completion time 6-8% shorter. Being retired leads to about 12% increase. Similarly, students and unemployed respondents are respectively 10-13% and 6-8% slower in completing a choice task. An additional explanation for these results is that on average retired people, students and unemployed individuals have a comparatively lower value of time because subject to less constrains. Those with an income greater than 50,000 spend approximately 10% less than lower income people. These might need more attention to be spent on evaluating price effects and their trade-offs with food attributes. However, when the number of income earners in the respondent's family increases so does τ_{nt}. More educated people tend to spend about 9-11% longer than those who did not graduate from high-school. We speculate that this might signal more attention to the task and the attributes, rather than an increased cognitive effort in completing the choice exercise. Three contextual variables concerning aspects of the respondent's place of residence are embedded within the model: (*i*) whether the respondent lives in an urban or rural area (*Urban*); (*ii*) the population size of the place of residence (*Population*); and (*iii*) a threshold time to walk from the respondent's house to open agricultural fields (*5 mins from fields*). We found that individuals living in urban or semi-urban areas spend about 8-9% longer than those living in rural or semi-rural areas. The coefficient for *Population* size is negative, providing further details about the size of the urban area: RT decreases as population size increases, ranging from 8.3% to 12.3%, implying that τ_{nt} is shorter in big urban conglomerates, where life moves at a fast pace. The positive sign of *5 mins from fields* chimes with this result, implying that τ_{nt} tends to be 2-4% longer for residents of locations close to agricultural fields, both in rural or urban areas. *Frequency HD sausage* indicates how many times, in the last year, the respondent has bought hot dog sausages. The observed sample values ranged between 0 and 80, with a median of 6. The higher this value the longer the completion time for the choice task (each purchase occasion increases time by 0.3-0.4% at the margin). High frequency buyers perhaps complete their choices more carefully because they are more intrigued by the innovative characteristics proposed in the DCE. A much stronger effect is that related to *Price HD sausage*: a unit increase in the price that the respondent considers appropriate for generic industrial HD sausages is related to a 23-36% increase in τ_{nt} at the margin. The variable *Hours from*

6am explores the potential decrease of cognitive energy from early morning (e.g., 6am is coded as 0, whereas 5am is coded as 23). Results indicate a positive relationship: all else being equal, the later in the day the survey was taken, the longer it took the respondent to complete the choice task. Although this effect obviously depends on the circadian rhythms and chronotypes of respondents (see Blatter & Cajochen, 2007), recent large-scale studies in Europe and China (Liu et al. 2020; Sládek et al. 2020) show that the chronotypes with eveningness or extreme eveningness are only around twenty percent of the population, with this proportion decreasing with age. *Mobile or tablet users* tend to spend about 8-9% longer in completing a choice task. This result is in line with previous research (Couper & Peterson, 2016; Liebe et al., 2015; Vass & Boeri, 2021; Wells et al., 2013) and suggests that comparing alternatives might be more difficult when using devices with smaller screen size. In addition, there are two possible explanations hypothesized by Gummer and Rossman (2015): first, the speed of the Internet connection might be slower for smartphones and tablets; and second, this type of devices is used outside home more often than computers, implying that these users may be subject to more distractions and interruptions during the survey. Finally, significant variation is explained by the market research firm (*Provider*) that provided the panel of respondents: those from firm 2, on average, have a completion time 10-11% shorter than those from firm 1. We speculate that this might be due to respondents in one market research firm being more experienced at DCE surveys than those in the other firm, who admitted not to be running DCE frequently.

From Table 2 emerges evidence of country-specific effects on the intercepts. These are expected given the results of tests on distribution equality.

Table 2 Log-linear panel regression with random effects for choice-task RT (minutes)

Dependent variable τ_{nt}	(1)	(2)	(3)
Independent variables	Model with respondent's characteristics and food habits	Model with respondent's characteristics, food habits and Shannon Index	Model with respondent's characteristics, food habits and Utility difference
Intercept	-1.791***	-1.915***	-1.489***
	(0.087)	(0.087)	(0.083)
Age 35-50	0.136***	0.137***	0.145***
	(0.022)	(0.022)	(0.022)
Age 50-65	0.317***	0.320***	0.353***
	(0.024)	(0.024)	(0.024)
Over 65	0.428***	0.387***	0.420***
	(0.051)	(0.048)	(0.046)
Man	-0.055***	-0.056***	-0.077***
	(0.017)	(0.017)	(0.017)
Retired	0.110***	0.111***	0.109***
	(0.039)	(0.039)	(0.037)
Student	0.120***	0.117***	0.091**
	(0.038)	(0.038)	(0.037)
Unemployed	0.070	0.073*	0.054
	(0.044)	(0.044)	(0.043)
Income > 50k	-0.093***	-0.091***	-0.091***
	(0.026)	(0.026)	(0.025)
Earners in family	0.032***	0.030***	0.022**
	(0.010)	(0.010)	(0.010)
Diploma or higher	0.106***	0.102***	0.089***
	(0.033)	(0.033)	(0.032)

Urban	0.074*** (0.023)	0.075*** (0.023)	0.083*** (0.022)
Population 50k-100k	-0.085*** (0.026)	-0.080*** (0.026)	-0.099*** (0.025)
Population 100k-500k	-0.083*** (0.026)	-0.081*** (0.026)	-0.102*** (0.025)
Population > 500k	-0.099*** (0.025)	-0.096*** (0.025)	-0.116*** (0.024)
5 mins from fields	0.034* (0.019)	0.032* (0.019)	0.018 (0.018)
Frequency HD sausage	0.004*** (0.001)	0.004*** (0.001)	0.003*** (0.001)
Price HD sausage	0.301*** (0.068)	0.305*** (0.068)	0.210*** (0.065)
Hours from 6am	0.002 (0.002)	0.003 (0.002)	0.003 (0.002)
Mobile or tablet user	0.082*** (0.025)	0.082*** (0.025)	0.077*** (0.024)
Provider	-0.097*** (0.038)	-0.105*** (0.037)	-0.107*** (0.036)
Italy	0.211*** (0.043)	0.267*** (0.047)	0.153*** (0.041)
United Kingdom	0.136*** (0.039)	0.116*** (0.039)	0.074* (0.038)
Italy x UK x Diploma or higher	-0.129*** (0.043)	-0.125*** (0.043)	-0.114*** (0.042)
UK x Over 65	-0.189*** (0.067)	-	-
Choice Task order	-0.086*** (0.001)	-0.086*** (0.001)	-0.083*** (0.001)
UK x Choice Task order	0.007*** (0.003)	0.007*** (0.003)	0.007*** (0.002)
H_{nt}	-	0.215*** (0.021)	-
$H_{nt} \times$ *Italy*	-	-0.104*** (0.031)	-
ΔV_{nt}	-	-	-0.011*** (0.002)
$opt.out_{nt}$	-	-	-0.425*** (0.010)
σ_η^2	0.419*** (0.006)	0.418*** (0.006)	0.408*** (0.006)
σ_w^2	0.484*** (0.004)	0.483*** (0.004)	0.460*** (0.004)
Observations	Unbalanced panel: n =2855, T = 1-10, N=27121		
Log-likelihood	-21785	-21710	-20487
Wald Chi-square	6229.86*** (df = 26)	6320.33*** (df = 27)	8248.88*** (df = 27)
Notes	*Robust standard errors in parentheses* *** p < 0.01; ** p < 0.05; * p < 0.1		

We find a positive and significant effect on RT for both the UK and Italy using Germany as a baseline. For example, respondents who graduated from high-school are generally slower, but those from the UK and Italy are faster on the margin, and so are the over 65 from the UK, but only in model (1), the one without any subjective measure of complexity. In all three models the variables describing the survey context also emerge as significant and with plausible signs: respondents using mobiles and tablets take longer, while choice tasks appearing later in the sequence require a shorter RT, which is a result in keeping with studies demonstrating a learning effect along the sequence of tasks (Day et al. 2012).

Model (2) has a positive and significant estimate for the coefficient of the Shannon index. So, the data fail to reject hypothesis 1. The estimated unit effect is 24% for the UK and Germany, but only 12% for Italy; Italians are well-known for being more customarily engaged in quality differentiation when choosing foods. Finally, Model (3) fits the data best, as shown by the increase in the log-likelihood value at the maximum. Hypotheses 2 and 3 cannot be rejected in this model. The expected negative effect of choice-task utility differences $\widehat{\Delta}V_{nt}$ on RT is very significant but small (1%), while the expected negative effect of the no-purchase option $opt.out_{nt}$ is also very significant and the largest in the entire regression (53%). The larger sample size and the larger set of explanatory variables accommodated in the panel regression allow for a more accurate estimation of the effect of $\widehat{\Delta}V_{nt}$, than in the model for τ_n reported in the Web Appendix (Table W.3). In that model the results did not support hypothesis 2. However, as Figure 5 demonstrates, the marginal effects forecasts come with much larger uncertainty for $\widehat{\Delta}V_{nt}$ than for $\widehat{H}_{nt}$. Translating the information of the forecast for $\widehat{H}_{nt}$ in operational practice for experimental design construction, one can expect a choice task with a $\widehat{H}_{nt}$ value of 0.4 to require about six seconds longer to be performed than one with a $\widehat{H}_{nt}$ value of 0.9. The uncertainty is much higher in the forecast for $\widehat{\Delta}V_{nt}$. The expected average RT in relation to other independent variables are presented in the Web Appendix, for the entire DCE sequence (Table W.4) and for the choice task (Table W.5).

Finally, both standard deviations for random effects are statistically significant across the three econometric models. This suggests that the random effects specification successfully captures the variations in RT; the one occurring at the choice task level as well as that across respondents.

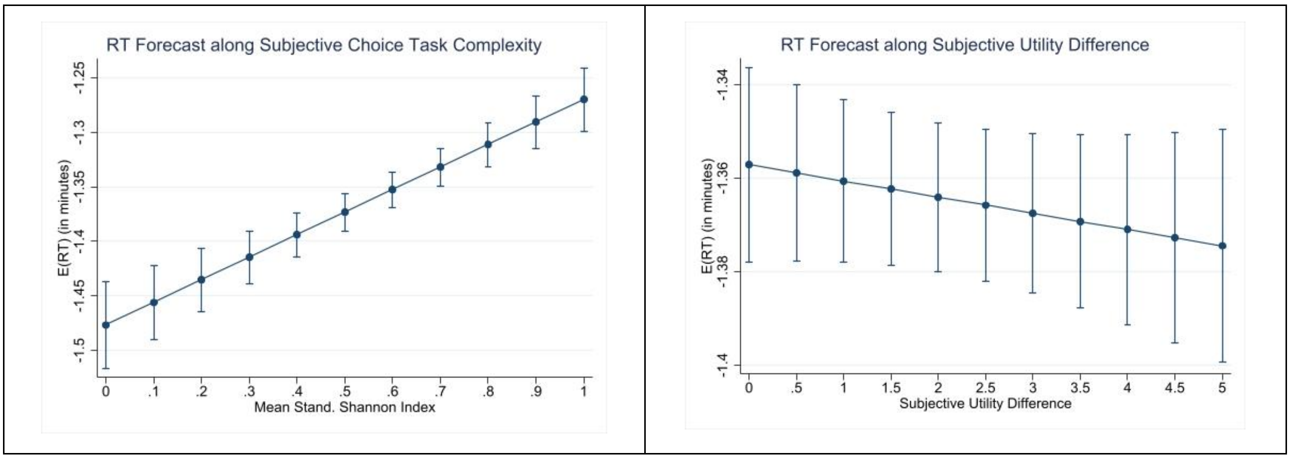

Figure 5 Forecast effects of choice complexity measures over their ranges.

6 Conclusions

This paper explores the determinants affecting response times (RT) by analyzing data collected from a web survey centered on a discrete choice experiment (DCE) with samples originating from three countries: Italy, Germany and the U.K. During the survey, respondents were asked to select one alternative from a set of three mutually exclusive options: either one of two hot dog sausages or a no-buy option. The choice exercise was repeated for ten choice tasks with focus on explaining the determinants of RT taken by respondents to complete the entire sequence of choice tasks in the DCE, as well as the time needed to answer each choice task. These RTs are interpreted as a crude manifestation of cognitive effort and are used as dependent variables in two separate log-linear regression models with which we empirically test our hypotheses. The underlying tenet is that if theoretical determinants of cognitive effort have the expected impact on RT, then we have prima-facie evidence that the two are meaningfully related. Hence, we use variables that proxy subjective cognitive resources, as well as measures of subjective and contextual choice complexity. These include sociodemographic, residential and economic characteristics of survey participants, various survey contextual factors and two measures of choice complexity; namely the Shannon Index and the absolute mean difference of utility of the food alternatives. Such measures were "subjectivised" by using the individual's estimates obtained from the estimation of a Mixed Multinomial Logit Model based on the observed choices.

The results of the panel log-linear model shows that respondents with low opportunity cost of time tend to spend longer on preferred choice selection and that RT decreased for choice tasks placed at later points in the sequence. Both Hypotheses 2 and 3, concerning other contextual and subjectivised measures of individual choice task complexity, fail to be rejected since the corresponding coefficients are statistically significant and with the expected negative signs. Overall, we conclude that RT, as interpreted as a proxy for cognitive effort, relate to key variables measuring contextual cognitive challenges and individual cognitive resources in a plausible fashion. Taken together the results corroborate the validity and quality of these types of preference data.

A number of limitations need to be considered. First, the link between RT and cognitive effort is an assumption based on previous research, but RT may not be the main indicator of cognitive burden and this relation may not hold true in all cases. In as much as we cannot observe the actual level of attention placed by respondents on the survey and cannot measure their cognitive effort, we are unable to distinguish between a respondent whose RT is due to having placed more attention and cognitive energy on the choice experiment and that of another who completed the survey while engaging in other tasks. In addition, the cognitive strategies used by respondents to complete a task may differ in terms of efficiency, leading to different levels of effort for similar RT. While this study offers a rich interpretation of the determinants of RT, our results cannot provide indications for a better design of choice experiments or surveys, considering that the relation between RT, cognitive effort and the quality of responses remains speculative. Future research could improve on this shortcoming, for example by analyzing the behavioral differences between fast and slow respondents for different web-based choice experiments, perhaps using neurological measurement of cognitive effort, as done in neuroeconomic approaches (Westbrook & Braver, 2015). Further, this study raised a question concerning the experimental design of choice experiments. Our findings show that response time--and, by direct correlation, cognitive effort--increase with subjectively evaluated choice task complexity, and that this is true across three countries inhabited by people with substantively different food cultures and traditions. Note that the current practice in DCE survey suggests that the selection of experimental designs ought to be based on statistical criteria of efficiency. This practice disregards the cognitive response of individual respondents to the complexity of choice tasks generated with such designs. We suggest that in using paradata from computer-based and web-administered choice experiments, a more targeted approach can be developed to better investigate this topic. This can significantly improve our understanding of the trade-offs between statistical efficiency and individual choice behavior, as suggested by Yao et al. (2015). To this effect one can envisage the development of adaptive designs in which, for example, early choices in the sequence of the web survey can probe the cognitive abilities of the respondents, to then adapt the degree of complexity of the subsequent set of choices in the later part of the sequence, as well as the expected time necessary to evaluate choice tasks of different complexity by respondents with given cognitive resources. This can be achieved by developing and implementing algorithms that can issue immediate feedback to respondents during surveys. For example, when abnormally long or short RTs are observed and these are unjustified by the complexity of choice tasks or by the cognitive resources available to the respondent, then immediate warning can be provided to respondents. These would invite them to reconsider the preferences expressed in the affected choices, and to avoid potential interruptions or distracting tasks in the rest of the survey (Höhne et al. 2017).

Alternatively, when abnormally long RTs are detected given the assessed complexity of the choice tasks, subsequent choice tasks can be reduced in terms of complexity by, for example, using alternatives that require fewer trade-offs (e.g. by keeping selected attribute levels fixed across alternatives). This would be an extension of the Conrad et al. (2017) approach to be tailored specifically to web-based DCE surveys. Computer-based survey technology can allow researchers to implement algorithms that can improve both statistical and cognitive efficiency and deliver a better experience to both the researcher and the respondent.

7 References

Arentze, T., Borgers, A., Timmermans, H., & Del Mistro, R. (2003). Transport stated choice responses: Effects of task complexity, presentation format and literacy, *Transportation Research Part E: Logistics and Transportation Review,* 39(3), 229-244. https://doi.org/10.1016/S1366-5545(02)00047-9

Bech, M., Kjaer, T., & Lauridsen, J. (2011). Does the number of choice sets matter? Results from a web survey applying a discrete choice experiment, *Health Economics,* 20, 273-286. https://doi.org/10.1002/hec.1587

Ben-Akiva, M., & Lerman, S.R. (1985). *Discrete Choice Analysis*. MIT Press, Cambridge, MA.

Blatter, H. & Cajochen, C. (2007). Circadian rhythms in cognitive performance: methodological constraints, protocols, theoretical underpinning. *Physiology and Behaviour*, 90, 196-208.

Bonsall, P., & Lythgoe B. (2009). Factors affecting the amount of effort expended in responding to questions in behavioural choice experiments. *Journal of Choice Modelling,* 2(2), 216-236. https://doi.org/10.1016/S1755-5345(13)70011-3

Börger, T. (2016). Are fast responses more random? Testing the effect of response time on scale in an online choice experiment, *Environmental and Resource Economics*, 65(2), 389-413. https://doi.org/10.1007/s10640-015-9905-1

Brown, T.C., Kingsley, D., Peterson, G.L., Flores, N.E., Clarke, A., & Birjulin, A. (2008). Reliability of individual valuations of public and private goods: choice consistency, response time, and preference refinement. *Journal of Public Economics*, 92(7), 1595-1606. https://doi.org/10.1016/j.jpubeco.2008.01.004

Callegaro, M., Villar, A., Krosnick, J., & Yeager, D. (2014). A critical review of studies investigating the quality of data obtained with online panels. In M. Callegaro, R. Baker, J. Bethlehem, A. Goritz, J. Krosnick & P. Lavrakas (Eds.), *Online panel research: A data quality perspective* (pp. 23-53). John Wiley & Sons.

Cameron, T.A., & DeShazo, J.R. (2010). Differential attention to attributes in utility-theoretic choice models. *Journal of Choice Modelling,* 3(3), 73-115. https://doi.org/10.1016/S1755-5345(13)70015-0

Campbell, D., Mørkbak, M.R., & Olsen, S.B. (2017). Response time in online stated choice experiments: the non-triviality of identifying fast and slow respondents. *Journal of Environmental Economics and Policy*, 6, 17-35. https://doi.org/10.1080/21606544.2016.1167632

Campbell, D., Mørkbak, M. R., & Olsen, S. B. (2018). The link between response time and preference, variance and processing heterogeneity in stated choice experiments. *Journal of Environmental Economics and Management*, 88, 18-34. https://doi.org/10.1016/j.jeem.2017.10.003

Campbell, D., & Erdem, S. (2019). Including opt-out options in discrete choice experiments: issues to consider. The Patient-Patient-Centered Outcomes Research, 12(1), 1–14. https://doi.org/10.1007/s40271-018-0324-6

Cantillo, V., & Ortúzar, J. de D. (2006). Implications of thresholds in discrete choice modelling. Transport Reviews, 26(6), 667–691. https://doi.org/10.1080/01441640500487275

Caspar, C.G. & Kroesen, M. (2014). On the (im-)possibility of deriving transport policy implications from hybrid choice models. *Transport Policy*, 36, 217-222. https://doi.org/10.1016/j.tranpol.2014.09.001

Caussade, S., Ortúzar, J. de D., Rizzi, L. I., & Hensher, D. A. (2005). Assessing the influence of design dimensions on stated choice experiment estimates. *Transportation research part B: Methodological*, 39(7), 621-640. https://doi.org/10.1016/j.trb.2004.07.006

Conrad, F.G., Couper, M.P., Tourangeau, R. & Zhang C. (2017). Reducing speeding in web surveys by providing immediate feedback. *Survey Research Methods,* 11(1), 45-61. https://doi.org/10.18148/srm/2017.v11i1.6304

Cook, J., Whittington, D., Canh, D.G., Johnson, F.R., & Nyamete, A. (2007). Reliability of stated preferences for cholera and typhoid vaccines with time to think in Hue, Vietnam. *Economic Inquiry,* 45, 100-114. https://doi.org/10.1111/j.1465-7295.2006.00038.x

Cook, J., Jeuland, M., Maskery, B., & Whittington, D. (2012). Giving stated preference respondents "time to think": Results from four countries. *Environmental and Resource Economics*, 51, 473-496. https://doi.org/10.1007/s10640-011-9508-4

Couper, M. (2000). Usability evaluation of computer-assisted survey instruments. *Social Science Computer Review*, 18(4), 384-396. https://doi.org/10.1177/089443930001800402

Couper, M., & Miller, P. V. (2008). Special issue: Web survey methods. *Public Opinion Quarterly,* 72(5), 831-835. https://doi.org/10.1093/poq/nfn066

Couper, M. P., & Peterson, G. J. (2016). Why Do Web Surveys Take Longer on Smartphones? Social Science Computer Review, 35(3), 357–377. https://doi.org/10.1177/0894439316629932

Crawford, S. D., Couper, M. P., & Lamias, M. J. (2001). Web surveys: Perceptions of burden. *Social Science Computer Review*, 19, 146–162. https://doi.org/10.1177/089443930101900202

DeShazo, J. R, & Fermo, F. (2002). Designing choice sets for stated preference methods: The effects of complexity on choice consistency. *Journal of Environmental Economics and Management,* 44(1), 123-143. https://doi.org/10.1006/jeem.2001.1199

Dhar, R. (1997). Consumer preference for a no-choice option. *Journal of Consumer Research*, 24(2), 215–231. https://doi.org/10.1086/209506

Day, B., Bateman I., Carson R., Dupont D., Louviere J., Morimoto S., Scarpa R., & Wang P. (2012). Ordering Effects and Choice Set Awareness in Repeat-Response Stated Preference Studies. *Journal of Environmental Economics and Management*, 63, 73-91. https://doi.org/10.1016/j.jeem.2011.09.001

Duquette, E.N. (2010). *Choice Difficulty and Risk Perceptions in Environmental Economics*. Ph.D. Thesis, Department of Economics, Graduate School - University of Oregon.

Espinoza-Varas, B., & Watson, C.S. (1994). Effects of decision criterion on response latencies of binary decisions. *Perception & Psychophysics,* 55(2), 190-203. https://doi.org/10.3758/BF03211666

Fan, W., & Yan, Z. (2010). Factors affecting response rates of the web survey: A systematic review. *Computers in Human Behavior,* 26(2), 132-139. https://doi.org/10.1016/j.chb.2009.10.015

Fazio, R. H. (1990). A practical guide to the use of response latency in social psychological research. In C. Hendrick & M.S. Clark (Eds.), *Research methods in personality and social psychology* (pp. 74-97). London, England: Sage.

Fischhoff, B. (2005). Cognitive processes in stated preference methods. In K.G. Maler & J. Vincent (Eds.), *Handbook of Environmental Economics* (pp. 937-968). Elsevier. https://doi.org/10.1016/S1574-0099(05)02018-8

Galesic, M., & Bosnjak, M. (2009). Effects of questionnaire length on participation and indicators of response quality in a web survey. Public opinion quarterly, 73(2), 349-360. https://doi.org/10.1093/poq/nfp031

Greszki, R., Meyer, M., & Schoen, H. (2015). Exploring the effects of removing "too fast" responses and respondents from web surveys. *The Public Opinion Quarterly*, 79(2), 471-503. https://doi.org/10.1093/poq/nfu058

Gummer, T. & Rossman, J. (2015). Explaining Interview Duration in Web Surveys: a Multilevel Approach, *Social Science Computer Review*, 33(2), 217-234 https://doi.org/10.1177/0894439314533479

Haaijer, R., Kamakura, W., & Wedel, M. (2000). Response latencies in the analysis of conjoint choice experiments. *Journal of Market Research*, 37(3), 376-382. https://doi.org/10.1509/jmkr.37.3.376.18775

Heerwegh, D. (2003). Explaining response latencies and changing answers using client-side paradata from a web survey. *Social Science Computer Review*, 21(3), 360-373. https://doi.org/10.1177/0894439303253985

Henninger, M., & Plieninger, H. (2020). Different styles, different times: how response times can inform our knowledge about the response process in rating scale measurement. *Assessment*, https://doi.org/10.1177/1073191119900003

Hess, S., & Stathopoulos, A. (2013). Linking response quality to survey engagement: A combined random scale and latent variable approach. *Journal of Choice Modelling*, 7, 1-12. https://doi.org/10.1016/j.jocm.2013.03.005

Hess, S., Rose, J. M., & Polak, J. (2010). Non-trading, lexicographic and inconsistent behaviour in stated choice data. *Transportation Research Part D: Transport and Environment*, 15(7), 405–417. https://doi.org/10.1016/j.trd.2010.04.008

Hillygus, D.S., Jackson, N., & Young, M. (2014). Professional respondents in non-probability online panels. In M. Callegaro, R. Baker, J. Bethlehem, A. Goritz, J. Krosnick & P. Lavrakas (Eds.), *Online panel research: A data quality perspective* (pp. 219-237). John Wiley & Sons. https://doi.org/10.1002/9781118763520.ch10

Höhne, J. K., Schlosser, S., & Krebs, D. (2017). Investigating cognitive effort and response quality of question formats in web surveys using paradata. *Field Methods*, 29(4), 365-382. https://doi.org/10.1177/1525822X17710640

Höhne, J. K., Schlosser, S., Couper, M. P., & Blom, A. G. (2020). Switching away: Exploring on-device media multitasking in web surveys. *Computers in Human Behavior*, 111, 106417. https://doi.org/10.1016/j.chb.2020.106417

Holmes, T., Alger, K., Zinkhan, C., & Mercer, E. (1998). The effect of response time on conjoint analysis estimates of rainforest protection values. *Journal of Forest Economics*, 4, 7-28.

Johnson, R.M., & Orme, B.K. (1996). How many questions should you ask in choice-based conjoint studies? Technical Paper, Sawtooth Software.

Johnston, R.J., Boyle, K.J., Adamowicz, W., Bennett, J., Brouwer, R., Cameron, T.A., Hanemann, W.M., Hanley, N., Ryan, M., Scarpa, R., Tourangeau, R., & Vossler, C.A., (2017). Contemporary guidance for stated preference studies. *Journal of the Association of Environmental and Resource Economics*, 4(2), 319-405. https://doi.org/10.1086/691697

Kaye-Blake, W. H. and Abell, W. L. & Zellman, E. (2009) Respondents' ignoring of attribute information in a choice modelling survey. *Australian Journal of Agricultural and Resource Economics*, 53(4), 547-564. https://doi.org/10.1111/j.1467-8489.2009.00467.x

Kreuter, F., Couper, M. P. & Lyberg, L. E. (2010*). The use of paradata to monitor and manage survey data collection.* Proceedings of the joint statistical meetings, pp. 282-296. American Statistical Association. Alexandria, VA, United States.

Kunz, T. & Gummer, T. (2019), Understanding Respondents' Attitudes Toward Web Paradata Use, *Social Science Computer Review*, 38(6), 739–753 https://doi.org/10.1177/0894439319826904

Lenzner, T., Kaczmirek, L., & Lenzner, A. (2010). Cognitive burden of survey questions and response times: A psycholinguistic experiment. *Applied Cognitive Psychology*, 24(7), 1003-1020. https://doi.org/10.1002/acp.1602

Liebe, U., Glenk, K., Oehlmann, M., & Meyerhoff, J. (2015). Does the use of mobile devices (tablets and smartphones) affect survey quality and choice behaviour in web surveys? *Journal of Choice Modelling*, 14, 17–31. https://doi.org/10.1016/j.jocm.2015.02.002

Lindhjem, H., & Navrud, S. (2011). Are Internet surveys an alternative to face-to-face interviews in contingent valuation? *Ecological Economics*, 70(9), 1628-1637. https://doi.org/10.1016/j.ecolecon.2011.04.002

Liu, Z., Dong, Y., & Xu, Y. (2020). Chronotype distribution in the Chinese population, *Brain Science Advances*, 6(2), 81-91. https://doi.org/10.26599/BSA.2020.9050004

Luce, R. D. (1957). *A Theory of Individual Choice Behavior.* Columbia University, Departments of Mathematical Statistics and Sociology and the Bureau of Applied Social Research.

Luce, R.D. (1986). *Response times: Their Role in Inferring Elementary Mental Organization*. New York: Oxford University Press.

MacMillan, D.C., Philip L., Hanley, N., & Alvarez-Farizo, B. (2002). Valuing the nonmarket benefits of wild goose conservation: A comparison of interview and group-based approaches. *Ecological Economics*, 43, 49-59. https://doi.org/10.1016/S0921-8009(02)00182-9

Matjašič, M., Vehovar, V. & Manfreda, K.L. (2018). Web survey paradata on response time outliers: A systematic literature review, *Metodološki zvezki*, 15(1), 23–41

Mayerl, J. (2013). Response latency measurement in surveys. Detecting strong attitudes and response effects. *Survey Methods: Insights from the Field*. https://doi.org/10.13094/SMIF-2013-00005

McClain, C. A., Couper, M. P., Hupp, A. L., Keusch, F., Peterson, G., Piskorowski, A. D., & West, B. T. (2019). A typology of web survey paradata for assessing total survey error. *Social Science Computer Review*, 37(2), 196-213. https://doi.org/10.1177/0894439318759670

McFadden, D. (1974). Conditional logit analysis of qualitative choice behavior. In P. Zarembka (Ed.), *Frontiers in Econometrics* (pp. 105-142). Academic Press.

McFadden, D. (1980). Econometric models for probabilistic choice among products. *Journal of Business*, 53(3), 13-29.

McFadden, D. (1999). Rationality for Economists? *Journal of Risk and Uncertainty*, 19(1–3), 73-105.

McFadden, D., & Train, K. (2000). Mixed MNL models for discrete response. *Journal of applied Econometrics*, 15(5), 447-470. https://doi.org/10.1002/1099-1255(200009/10)15:5<447::AID-JAE570>3.0.CO;2-1

Merten, T. & Ruch, W. (1996). A comparison of computerised and conventional administration of the German version of the Eysenck personality questionnaire and the Carnell rating scale for depression. *Personality and Individual Differences*, 20(3), 281-291. https://doi.org/10.1016/0191-8869(95)00185-9

Revelt, D., & Train, K. (1998). Mixed logit with repeated choices: Households' choices of appliance efficiency level. *Review of Economics and Statistics*, 80(4), 647-657. https://doi.org/10.1162/003465398557735

Rose, J. & Black, I. (2006). Means matter, but variance matter too: Decomposing response latency influences on variance heterogeneity in stated preference experiments. *Marketing Letters*, 17, 295-310. https://doi.org/10.1007/s11002-006-8632-3

Safir, A., & Black, T., & Steinbach, R. (2001). *Using paradata to examine the effects of interviewer characteristics on survey response and data quality.* Proceedings of the Annual Meeting of the American Statistical Association. Atlanta, GA, United States.

Scarpa, R. & Rose, J.M. (2008). Design efficiency for nonmarket valuation with choice modelling: How to measure it, what to report and why. *Australian Journal of Agricultural and Resource Economics*, 52(3), 253-282. https://doi.org/10.1111/j.1467-8489.2007.00436.x

Scarpa, R., Gilbride, T. J., Campbell, D., & Hensher, D. A. (2009). Modelling attribute non-attendance in choice experiments for rural landscape valuation. *European Review of Agricultural Economics*, 36(2), 151–174. https://doi.org/10.1093/erae/jbp012

Scarpa, R., Notaro, S., Louviere, J.J. & Raffaelli, R. (2011). Exploring Scale Effects of Best/Worst Rank Ordered Choice Data to Estimate Benefits of Tourism in Alpine Grazing Commons, *American Journal of Agricultural Economics*, 3(3), 813-828. https://doi.org/10.1093/ajae/aaq174

Scott, A., Jeon, S.H., Joyce, C.M., Humphreys, J.S., Kalb, G., Witt, J., & Leahy, A. (2011). A randomised trial and economic evaluation of the effect of response mode on response rate, response bias, and item non-response in a survey of doctors. *BMC Medical Research Methodology*, 11, 1-12. https://doi.org/10.1186/1471-2288-11-126

Sládek, M., Kudrnáčová Röschová, M., Adámková, V., Hamplová, D., & Sumová, A. (2020). Chronotype assessment via a large scale socio-demographic survey favours yearlong Standard time over Daylight Saving Time in central Europe. *Scientific Reports (Nature Research)* 10, 1419 https://doi.org/10.1038/s41598-020-58413-9

Stern, M. J. (2008). The use of client-side paradata in analyzing the effects of visual layout on changing responses in web surveys. *Field Methods*, 20(4), 377-398. https://doi.org/10.1177/1525822X08320421

Svedsäter, H. (2007). Ambivalent statements in contingent valuation studies: Inclusive response formats and giving respondents time to think. *Australian Journal of Agricultural and Resource Economics*, 51, 91-107. https://doi.org/10.1111/j.1467-8489.2007.00346.x

Swait, J. (2001). A non-compensatory choice model incorporating attribute cutoffs. Transportation Research Part B: Methodological, 35(10), 903–928. https://doi.org/10.1016/S0191-2615(00)00030-8

Swait, J., & Adamowicz, W. (2001a). Choice environment, market complexity, and consumer behavior: A theoretical and empirical approach for incorporating decision complexity into models of consumer choice. *Organizational Behavior and Human Decision Processes*, 86(2), 141-167. https://doi.org/10.1006/obhd.2000.2941

Swait, J., & Adamowicz, W. (2001b). The influence of task complexity on consumer choice: A latent class model of decision strategy switching. *Journal of Consumer Research*, 28(1), 135-148. https://doi.org/10.1086/321952

Train, K. E. (2009). *Discrete choice methods with simulation*. 2nd edition. Cambridge University Press.

Tversky, A. (1969). Intransitivity of preferences. Psychological Review, 76(1), 31. https://doi.org/10.1037/h0026750

Vass, C. M., & Boeri, M. (2021). Mobilising the Next Generation of Stated-Preference Studies: the Association of Access Device with Choice Behaviour and Data Quality. *The Patient-Patient-Centered Outcomes Research*, 14(1), 55–63. https://doi.org/10.1007/s40271-020-00484-x

Vista, A.B., Rosenberger, R.S., & Collins, A.R. (2009). If you provide it, will they read it? Response time effects in a choice experiment. *Canadian Journal of Agricultural Economics*, 57(3), 365–377. https://doi.org/10.1111/j.1744-7976.2009.01156.x

Vij, A. & Walker, J. L. (2016). How, when and why integrated choice and latent variable models are latently useful. *Transportation Research Part B: Methodological*, 90, 192-217. https://doi.org/10.1016/j.trb.2016.04.021

Vörös, Z., & Rouet, J-F. (2016). Laypersons' digital problem solving: Relationships between strategy and performance in a large-scale international survey. *Computers in Human Behavior*, 64, 108-116. https://doi.org/10.1016/j.chb.2016.06.018

Wells, T., Bailey, J. T., & Link, M. W. (2013). Comparison of Smartphone and Online Computer Survey Administration. *Social Science Computer Review*, 32(2), 238–255. https://doi.org/10.1177/0894439313505829

Westbrook, A., & Braver, T.S. (2015). Cognitive effort: A neuroeconomic approach. *Cognitive, Affective, & Behavioral Neuroscience* 15(2), 395–415. doi: 10.3758/s13415-015-0334-y

Whittington, D., Smith, V.K., Okorafor, A., Okore, A., Liu, J.L., & McPhail, A. (1992). Giving respondents time to think in contingent valuation studies: A developing country application. *Journal of Environmental Economics and Management*, 22(3), 205-225. https://doi.org/10.1016/0095-0696(92)90029-V

Yan, T., & Tourangeau, R. (2008). Fast times and easy questions: The effects of age, experience and question complexity on web survey response times. *Applied Cognitive Psychology*, 22, 51-68. https://doi.org/10.1002/acp.1331

Yao, R.T., Scarpa, R., Rose, J.M. & Turner, J.A. (2015). Experimental design criteria and their behavioural efficiency: An evaluation in the field. *Environmental and Resource Economics*, 62, 433-455. https://doi.org/10.1007/s10640-014-9823-7

Zhang, C. & Conrad, F. (2014). Speeding in web surveys: The tendency to answer very fast and its association with straightlining. *Survey Research Methods*, 8(2), 127-135.