European Review of Agricultural Economics Vol **41 (4)** (2014) pp. 627–655
doi:10.1093/erae/jbt035
Advance Access Publication 3 December 2013

Distinguishing beliefs from preferences in food choice

Jayson L. Lusk*, Ted C. Schroeder and Glynn T. Tonsor
Kansas State University, Manhattan, KS, USA

Received October 2012; final version accepted September 2013

Review coordinated by Iain Fraser

Abstract

In the past two decades, there has been an explosion of studies eliciting consumer willingness-to-pay for food attributes; however, this work has largely refrained from drawing a distinction between preferences for health, safety and quality on the one hand and consumers' subjective beliefs that the products studied possess these attributes, on the other. Using data from three experimental studies, along with structural economic models, we show that controlling for subjective beliefs can substantively alter the interpretation of results and the ultimate implications derived from a study. The results suggest the need to measure subjective beliefs in studies of consumer choice and to utilise the measures when making policy and marketing recommendations.

Keywords: beliefs, rank-dependent expected utility, willingness-to-pay

JEL classification: Q13, Q18, C91, D83

> How a person chooses among potential alternatives is not only a matter of 'what he wants' but also of 'what he believes,' and for some kinds of choices an actor's beliefs . . . may play a most crucial role.
>
> – James Buchanan (1991, pp. 52–53)

1. Introduction

Economists conduct hundreds of studies each year eliciting consumers' willingness-to-pay (WTP) for various food, health and environmental outcomes (e.g. see Adamowicz, 2004; Dannenberg, 2009; Grunert *et al.*, 2009; Lagerkvist and Hess, 2011). The WTP values are used to inform cost–benefit analyses, improve firm-level marketing decisions and to better understand the nature of consumer choice. However, this large body of applied work often fails to explicitly acknowledge the fact that WTP estimates are composed of a combination of preferences and beliefs, which are the subjective probabilities of attaining different

*Corresponding author. Oklahoma State University, Stillwater, OK, USA. E-mail: jayson.lusk@okstate.edu

outcomes.[1] As we will show, failing to distinguish between these two determinants of WTP can lead to serious misinterpretations of a study's findings.

Separating beliefs from preferences is important for a number of reasons. First, economists traditionally view preferences as relatively stable constructs while beliefs are more malleable; thus, understanding changes in WTP, which are often conceptualised as occurring through a Bayesian updating process, requires recognition of the heterogeneity and evolution of beliefs (Lusk *et al.*, 2004; Pennings and Wansink, 2004; Huffman *et al.*, 2007). Second, belief elicitation is needed because people do not always equally believe the 'objective' information on probabilities presented in advertisements, experiments or surveys (e.g. Hayes *et al.*, 1995). Teisl and Roe (2010), for example, show that people's perceptions of the likelihood of getting sick from food-borne illness can differ from 'objective' probabilities of food contamination, in part due to the ability for self-protection through cooking. Third, not only are beliefs subjective, a large body of psychology literature (e.g. Kahneman and Tversky, 2000) suggests people distort beliefs (or the probabilities of attaining certain outcomes). As such, one cannot simply determine the welfare effects of a policy by *ex post* applying 'objective' probabilities of uncertain outcomes to WTP values measured in a survey or experiment (see the discussion in Roberts, Boyer and Lusk, 2008). As Leggett (2002) puts it (p. 343), 'Because these perceptions – rather than objective, scientific measurements of quality – are what ultimately determine choices, standard welfare estimates derived from these choices will be incorrect when perceptions are wrong.'

Econometric approaches that do not account for differences in beliefs across people may yield misleading estimates of welfare changes from a policy (Marette, Roe and Teisl, 2012). Leggett (2002), for example, argued that when objective probability or quality measures are used to estimate a model instead of people's subjective beliefs (p. 344), 'the parameters of the preference function estimated by the researcher using data on [objective or true quality] will be incorrect . . . the researcher will be able to recover the parameters of the preference function only by obtaining data on [perceptions or beliefs] through surveys or by estimating [perceptions or beliefs] using a model of perception formation.' Welfare analysis also requires knowledge of people's beliefs about the status quo. For example, introduction of 'objective' food labels can create uncertainty and change beliefs about the quality of unlabelled products, which complicates welfare analysis (e.g. Dannenberg, Scatasta and Strum, 2011).

Ultimately, measures of beliefs are needed to *understand* why consumers make the choices they do. Manski (2004), for example, noted that, in the widely studied ultimatum game, a proposer may decide to split the pie either because they care about their partner or because they believe their partner will reject an unequal offer; both possibilities can rationalise the frequently

1 We are not, of course, arguing that the issue has not been studied as the forgoing literature review will reveal. However, these developments have largely occurred outside the context of studies focused on food policy and consumer WTP for food.

observed 50–50 split, but, without measuring beliefs, one cannot discriminate between the two competing hypotheses.

To make these arguments more concrete, consider a consumers' choice of whether to buy organic food. A decision *not* to buy organic could arise because the consumer does not sufficiently value environmental or health outcomes to justify the price premium, or it could be that the consumer does not believe organics are healthier or are better for the environment. If the latter is the primary driver of choice, then a marketer might offer alternative products or production systems with greater credibility (e.g. including third party verification) or they might try to modify existing beliefs via advertising. If the former is the primary driver of choice, then it might be more fruitful to pay attention to factors such as income and relative prices. Of course, these issues are interrelated, and choice is driven by a combination of preferences and beliefs; however, the general point remains.

In a sense, it seems almost trivial to suggest that 'beliefs matter' and should be elicited in valuation studies. Yet, the fact that belief elicitation is so uncommon makes the topic worthy of further exploration. As Manski (2004, p. 1330) put it, 'The prevailing practice has been to assume that decision makers have specific expectations that are objectively correct (i.e. rational). This practice reduces the task of empirical inference to revelation of preferences alone, but has contributed to a crisis of credibility . . . . I have concluded that econometric analysis of decision making with partial information cannot prosper on choice data alone. However, combination of choice data with other data should mitigate the credibility problem and improve our ability to predict behavior. The data I have in mind are self-reports of expectations elicited in the form called for by modern economic theory; that is, subjective probabilities.' This paper attempts to put Manski's suggestion into practice.

Using data from three illustrative studies, the purpose of our paper is to clearly outline some approaches that future researchers can use to analyse beliefs and preferences, and hopefully spur additional, subsequent research on the topic. In three separate studies, we show how to couple choice or WTP data with beliefs measured in post-experiment surveys and reveal that the merger of such data yields insights and inferences that could not otherwise have been obtained. Study 1 utilises data from a non-hypothetical choice experiment conducted with beef steaks and reveals that choices between steaks are significantly related to people's beliefs about the extent to which the steaks are safe and tender. Study 2 shows that the findings from study 1 hold up when the context is moved from a choice experiment to an experimental auction; we find that WTP for different beef steaks is significantly related to people's beliefs about the extent to which the steaks are safe and tender. In both cases, we find that the implied values of safety and tenderness significantly change once one accounts for beliefs. In the last study, we utilise data from a non-hypothetical choice experiment to show that consumers' WTP for meat with country-of-origin labels is significantly related to their beliefs about the origin of unlabelled products. The following section provides background on the topic, then the next three sections discuss each of the empirical studies and the last section concludes.

2. Background

The standard model employed in economics to determine how consumers assess the desirability of a choice option with uncertain outcomes (or attributes) is expected utility theory (EUT), which was formalised by von Neumann and Morgenstern (1944). Under EUT, individual *i* evaluates risky prospect *j* as follows:

$$\mathrm{EU}_{ij} = \sum_{k=1}^{K} p_{ijk} U(x_k), \tag{1}$$

where p_{ijk} is the probability of individual *i* receiving outcome (or attribute) x_k from option *j* and $U(\cdot)$ is a utility function that describes the desirability of attaining the outcomes/attributes. Although x_k is typically interpreted as a dollar amount, it is also appropriate to interpret x_k as a variable indicating the presence/absence of a discrete attribute (e.g. organic, local, etc.) or a continuous quantity of some attribute (e.g. fat content, sodium content, etc.), given a sufficiently general utility function $U(\cdot)$. Although most WTP studies are constructed such that attributes are assumed to be known with certainty, i.e. $p_{ijk} = 1$, there are a number of examples where uncertainty is explicitly modelled (e.g. Hayes *et al.*, 1995; Teisl and Roe, 2010).

In the EUT formulation, the probabilities, p_{ijk}, are often taken to be objective facts, such as the probability of a coin toss turning up heads. However, in most real-world applications, objective probabilities are unknown, and as such, the probabilities in Equation (1), p_{ijk}, are typically subjective and individual specific, as indicated by the *i* subscript. Thus, p_{ijk} are also known as *beliefs*.

In this subjective expected utility (SEU) framework, formally developed by Savage (1954), the utility a consumer expects to derive from a product is composed of two components: the desire to obtain the outcomes provided by the product, given by $U(x_k)$ in Equation (1), *and* their subjective beliefs that the product or policy will actually deliver the outcomes, given by p_{ijk} in Equation (1). Although economists and psychologists have widely acknowledged this fact for more than half a century, its implications have not been widely incorporated in consumer WTP studies focused on food policy and marketing.[2]

For example, in choice data analyses, the selection of option A over B is often interpreted as indicating a preference of A over B. This is true, but the traditional notion of 'preference' used in many welfare analyses relates only to the utility consumers derive from outcomes (i.e. health, safety, etc.) that option A provides that B does not, i.e. 'preference' relates to the function $U(x_k)$ in Equation (1). However, to isolate this 'preference', one must also know the extent to which the consumer *believes* option A to possess more

2 One distinction drawn by many psychologists and some economists (e.g. Kahneman and Sugden, 2005) is between experience and decision utility. These authors argue for experience utility, as measured, for example, by happiness or life satisfaction scales as a basis for policy evaluation. Our definition of utility used here is that of decision utility, and measured beliefs relate to those that affect choices people make. We follow the standard approach in economics when using these decision utilities to calculate a welfare measure like willingness-to-pay.

health, safety, etc. than option B. Stated differently, most WTP studies confound beliefs and preferences (e.g. Lusk and Schroeder, 2004; Loureiro and and Umberger, 2007).

More concretely, note that a choice of option A over option B reveals that $\mathrm{SEU}_{i\mathrm{A}} > \mathrm{SEU}_{i\mathrm{B}}$ or $\sum_{k=1}^{K} p_{i\mathrm{A}k}U(x_k) > \sum_{k=1}^{K} p_{i\mathrm{B}k}U(x_k)$, or re-arranging, $\sum_{k=1}^{K} (p_{i\mathrm{A}k} - p_{i\mathrm{B}k})U(x_k) > 0$. Thus, a choice of A over B reveals something about the desirability of the outcomes provided, $U(x_k)$, but the choice also reflects differences in the perceptions (or beliefs) that A and B will actually yield a given outcome $(p_{i\mathrm{A}k} - p_{i\mathrm{B}k})$. Many consumer studies either ignore beliefs altogether and report whether people prefer, say, outcome $k = 1$ to outcome $k = 2$, i.e. whether $U(x_1) > U(x_2)$, and the WTP to have x_1 rather than x_2, or they confound beliefs and preferences and simply report whether $\mathrm{SEU}_{i\mathrm{A}} > \mathrm{SEU}_{i\mathrm{B}}$ and the WTP to have A vs. B.

There are two broad problems with these typical approaches; the first relates to the interpretation and implication of results and the second deals with proper measurement of the effects of interest. On the issue of interpretation, suppose all that is known is consumer WTP to have outcome x_1 rather than x_2. For example, suppose that one only knows consumer WTP to have non-genetically modified (non-GM) food rather than GM food. A policy analyst might take these WTP figures to extrapolate the implications of a ban on GM. However, even in the presence of a ban, some people would probably still *believe* some food is GM either due to misinformation, perceptions about imperfect enforceability of the law or malfeasance by food companies. Thus, the 'true' welfare gains from such a policy are not equal to the WTP to have non-GM vs. GM food but rather the welfare benefits of the policy must be discounted by the consumers' beliefs that post-ban food is, in fact, GM.

It might be tempting to dismiss issues like those described in the previous example as easily surmounted by applying 'objective' probabilities or beliefs elicited in other contexts to WTP elicited in prior studies, but this would ignore the challenge of measurement. In particular, if beliefs are not utilised during the estimation of WTP to have outcome x_1 rather than x_2, then it is likely that the WTP estimates will be biased because of the belief-preference confound, which has not been accounted for in the estimation. The analyst derives estimates by assuming choices were made based on objective probabilities or qualities, when in fact they were being driven by subjective beliefs. Even in 'simple' surveys or experiments, where people are asked how much they are WTP, for example, for non-GM vs. GM, it cannot be guaranteed that the subjects believe the researchers that non-GM products are 100 per cent non-GM or that GM are actually GM (especially in countries where it is not currently produced).

Although the current paper works within the SEU framework and uses the term 'beliefs' to refer to subjective probabilities, p_{ijk}, it is useful to consider other related approaches stemming from psychology and marketing. One of the most widely cited models in psychology and marketing is the theory of reasoned action developed by Fishbein and Ajzen (1975). Their model posits that an individual's propensity to undertake a behaviour (such as purchasing a product) is driven by the person's attitude towards the behaviour. The attitude, in turn, is driven by beliefs that the behaviour will generate a given outcome multiplied by an evaluation of the beliefs (i.e. whether and to what extent the behaviour will lead to an outcome deemed desirable or undesirable). The model is similar to the SEU approach in that a behaviour (or choice) is assumed to be driven by sum of beliefs multiplied by 'preferences' or 'evaluations'. Other approaches, like the total food quality model of Grunert (2005) do not mention beliefs explicitly, but rather utilise related constructs such as perceived quality and perceived safety, which relate to subjective expectations that particular quality or safety outcomes will be obtained. However, empirical studies based on the theory of reasoned action or the total food quality model almost always measure beliefs or perceptions using Likert-type scales and behavioural intentions are measured in ways that often prohibit the use of economic welfare analysis.

Our paper is by no means the first to attempt to measure or utilise beliefs. Within the SEU framework, subjective beliefs have been studied in the context of development (e.g. Delavande *et al.*, 2011; Bellemare, 2012) and environmental and health economics (e.g. Cameron, 2005; Cameron, DeShazo and Johnson, 2011). In addition, the *Journal of Applied Econometrics* recently published an issue on measurement and analysis of beliefs (Bellemare and Manski, 2011). However, few of these studies have focused on food and agricultural issues. In the studies that have focused on food policy or marketing, the role of beliefs vs. preferences/attitudes have primarily been studied using Likert-type scale questions without specifically estimating WTP or policy impacts (e.g. see Lusk and Coble, 2005; Pennings, Wansink and Meulenberg, 2002; Schroeder *et al.*, 2007). The analyses in Teisl and Roe (2010) and Marette, Roe and Teisl (2012) are more similar to that employed here in the sense that they explicitly incorporate probabilistic beliefs in their analysis of consumer choice.

It is also worthwhile to briefly mention the large literature on methods of belief elicitation. We have already mentioned approaches using psychometric scales which utilise questions such as: 'On a scale if 1 to 5, where 1 is very unlikely and 5 is very likely, how likely do you think it is that . . . .' As discussed by Viscusi and Hakes (2003), these and other scale-type questions do not measure probabilities. More germane to the present study are approaches designed to elicit subjective probabilities. Within the economics literature, numerous theoretical and empirical studies have analysed the properties of belief elicitation methods designed to provide incentives for 'truthful' revelation. Typically, these elicitation mechanisms involve some sort of quadratic or logarithmic scoring rule where individuals are rewarded for good predictions and penalised for poor predictions (Savage, 1971; Holt, 1986; Karni, 2009). One challenge with these approaches is that they often require questionable assumptions (such as risk neutrality) to be incentive compatible, but even more problematic is that knowledge of the 'true' or 'objective' probabilities needs be known to determine pay-offs. Prelec (2004) proposed a scoring rule that provides incentives for truthful revelation. His mechanism is noteworthy in that it does not rely on an objective estimate of the outcome of interest; however, his method is quite complicated and it is unclear how useful it might ultimately be in empirical applications. More generally, there are a large number of methods that have been previously used to elicit subjective probabilities (or probability distributions) including direct numerical elicitation (which is the method we used here), use of probability wheels or bars, bisection methods that ask subjects to subdivide a range of values into intervals that are equally likely and indirect methods that infer beliefs from choices, rating and rankings (for comparisons of these methods, see Hora, Hora and Dodd, 1992; O'Hagan *et al.*, 2006; Spetzler and von Holstein, 1975 among others).

3. Study 1: effects of beliefs in a non-hypothetical choice experiment

In the first study, we utilise data from a non-hypothetical choice experiment regarding beef steaks and merge it with data collected in a post-experiment survey that elicited beliefs regarding the likelihood that competing steaks possessed various safety and quality attributes.

3.1. Procedure

The non-hypothetical choice data come from the study on hypothetical bias conducted by Lusk and Schroeder (2004).[3] In their study, randomly recruited subjects from Manhattan, KS were asked to make 17 discrete choices. In each choice, subjects had to choose between five types of steaks [generic – unlabelled, guaranteed tender, 'natural' (i.e. no growth hormones or antibiotics), Choice (a USDA grade of beef), or Certified Angus Beef (CAB)], or a 'no purchase' option. An information sheet was distributed describing each of these steak types and each of the steaks could be visually appraised as well. Across the 17 choices, the prices of each steak type varied in such a way that prices were uncorrelated with steak type. Once all 17 choices were completed, one was randomly selected as binding and subjects purchased the steak type chosen in the binding task (if 'no purchase' was selected, the subject made no purchase).

Following the completion of the 17 choice tasks, subjects completed a post-experiment survey. In the survey, standard questions about gender, age, education and income were asked. In addition, we asked questions about the subjects' beliefs regarding the steak types used in the choice experiment (note: these data have not previously been analysed or reported). The survey queried subjects' beliefs about the tenderness of each steak type and the perceived safety of the generic steak. We asked three questions of the form:

> If you were to purchase a ≪*type*≫ steak, what is the likelihood that this steak would be tender? For example, a 0% chance would mean there is NO chance the ≪*type*≫ steak would actually be tender; whereas, a 100% chance would mean that the ≪*type*≫ steak

3 We have also conducted the following analysis using the hypothetical choice data reported in Lusk and Schroeder (2004), and the results are broadly consistent with that from the non-hypothetical data in terms of the effects of beliefs on choice.

would be tender for certain. There is a _______% chance the ≪*type*≫ steak will be tender.

where ≪*type*≫ was either generic, Choice quality grade or CAB. We did not ask subjects' beliefs about the tenderness of the guaranteed tender steak or the natural steak. In addition, we also asked about the perceived health effects of the generic steak resulting from the use of growth hormones and antibiotics:

> If you were to purchase a *generic* steak, what is the likelihood that you, at some point in the future, will become ill due to the possible use of added growth hormones or antibiotics? For example, a 0% chance would mean there is NO chance you will become ill due to possible use of added growth hormones or antibiotics in the *generic* steak; whereas, a 100% chance would mean that you will definitely become ill at some point in the future due to use of added growth hormones and antibiotics in the *generic* steak. There is a ______% chance I will become ill at some point in the future due to the possible use of added growth hormones or antibiotics in the *generic* steak.

We make use of data from 37 individuals who completely answered the entire post-experiment survey, each of whom answered 17 choice questions, yielding a data set of $37 \times 17 = 629$ choices.

3.2. Model

Aside from the non-hypothetical nature of the decision task, the data collection approach was a version of the oft-used choice experiment method (e.g. Louviere, Hensher and Swait, 2000). Analysis of such data would typically proceed by estimating a random utility model (McFadden, 1973) in which a preference parameter is estimated for each attribute (in this case, each steak type) and for price. In what follows, we show how the choice data can be augmented with data on beliefs in a theoretical consistent way.

Let the random *expected* utility of steak option j for individual i be given by:

$$\mathrm{EU}_{ij} = \gamma_j + (1 - P_{ij}^{\mathrm{ill}})U(\mathrm{well}) + P_{ij}^{\mathrm{tender}}U(\mathrm{tender}) - \alpha\mathrm{Price}_j + \varepsilon_{ij}, \quad (2)$$

where P_{ij}^{ill} is subject i's belief that steak type j will cause illness due to the use of growth hormones and antibiotics; $U(\mathrm{well})$ is the relative preference for wellness over illness (note: the preference for illness, $U(\mathrm{ill})$, has implicitly been normalised to zero for identification such that $U(\mathrm{well})$ is the difference in utilities of wellness and illness: $U(\mathrm{well}) - U(\mathrm{ill})$); P_{ij}^{tender} is subject is belief that steak type j will be tender; $U(\mathrm{tender})$ is the relative preference for steak tenderness over toughness (note: the preference for toughness has been implicitly normalised to zero for identification such that $U(\mathrm{tender})$ is the difference in utilities of tender and tough: $U(\mathrm{tender}) - U(\mathrm{tough})$; Price_j is the price of steak type j; α is the marginal utility of income; ε_{ij} is a stochastic error term; and γ_j is an 'alternative specific constant', which reflects preferences for steak type j that are not accounted for by beliefs about and preferences for tenderness, safety and price.

The difference between Equation (2) and the typical model estimated in choice experiment studies is the inclusion of the belief variables. For example, the 'reduced form' indirect utility function (Lusk and Schroeder, 2004) estimated was $\tilde{\gamma}_j - \tilde{\alpha}\text{Price}_j + \tilde{\varepsilon}_{ij}$. The implication is that $\tilde{\gamma}_j$ in the reduced form model implicitly incorporated beliefs about the relative tenderness and safety of steak type j. As such, the typical approach cannot isolate the relative contributions of tenderness and safety concerns to the overall preference for steak type j. With Equation (2), the consumers' WTP for tenderness is given by $U(\text{tender})/\alpha$; however, no similar insight is obtainable from the reduced form estimates. Even comparing $\tilde{\gamma}_{\text{guaranteed tender}}$ to $\tilde{\gamma}_{\text{generic}}$ in the reduced form model does not yield the complete value for tenderness unless people believe there is absolutely no chance the generic steak is tender (an assumption not supported by the data). Moreover, $\tilde{\gamma}_j$ is likely a biased estimate of the 'true' expected quality, $\gamma_j + (1 - P_{ij}^{\text{ill}})U(\text{well}) + P_{ij}^{\text{tender}}U(\text{tender})$, given that the latter involves heterogeneity in beliefs across people that are likely correlated with the error term, $\tilde{\varepsilon}_{ij}$.

An additional consideration is that psychological research shows that the expected utility model shown in Equation (2) is not always a good descriptive model of consumer choice. In particular, evidence suggests individuals tend to under-weight low-probability events and over-weight high-probability events (although a few studies have found the opposite, e.g. Birnbaum and Chavez, 1997). To account for such behaviour, we utilise Quiggin's (1982) rank-dependent model, which was subsequently incorporated into Tversky and Kahneman's (1992) cumulative prospect theory. Under this approach, the random rank-dependent expected utility of steak option j is

$$\text{RDEU}_{ij} = \gamma_j + w(1 - P_{ij}^{\text{ill}})U(\text{well}) + w(P_{ij}^{\text{tender}})U(\text{tender}) - \alpha\text{Price}_j + \varepsilon_{ij}, \tag{3}$$

where $w(P)$ is a probability weighting function, which we assume takes the form proposed by Tversky and Kahneman (1992): $w(P) = P^{\delta}/(P^{\delta} + (1-P)^{\delta})^{1/\delta}$. When $\delta = 1$, people weight probabilities linearly and Equation (3) collapses back to Equation (1). Previous estimates of δ have fallen in the range of 0.56–0.71 (e.g. see Camerer and Ho, 1994; Tversky and Kahneman, 1992; Wu and Gonzalez, 1996). Although many studies have found evidence of probability weighting, such a model does not always describe the behaviour of all subjects (Hey and Orme, 1994; see also Shleifer, 2012 for some critiques of prospect theory).

Assuming the error terms in Equation (2) [or Equation (3) depending on the model estimated] are distributed iid extreme value type I, the conventional multinomial logit model is obtained, and the parameters, γ_j, $U(\text{well})$, $U(\text{tender})$ and α can be estimated via conventional maximum likelihood estimation. Rather than proceeding with the multinomial logit, however, we estimated a random-parameter logit (RPL) specification because of the repeated nature of the data (i.e. each person answered 17 choice questions) and because of the potential heterogeneity in preferences. More specifically, we specified the generic steak parameter for individual i as normally distributed with mean $\overline{\gamma_{\text{generic}}}$ and standard deviation σ_{generic}, i.e., $\gamma_{\text{generic},i} = \overline{\gamma_{\text{generic}}} + e_i\sigma_{\text{generic}}$, where $e_i \sim N(0, 1)$. We also consider specifications where other model variables are similarly specified as random.

Because we did not ask safety and tenderness belief questions for all steak types, a few assumptions had to be made in order to estimate the model. First, we assume that people believe the guaranteed tender steak to be tender with certainty, i.e. $P^{\text{tender}}_{\text{guaranteed tender}} = 1$, and that the natural steak will not cause illness due to growth hormones and antibiotics, i.e. $P^{\text{ill}}_{\text{natural}} = 0$. We did not ask subjects about their beliefs regarding the perceived health consequences of growth hormones and antibiotics for the guaranteed tender, Choice or CAB steaks because they did not differ from the generic steak in this regard. As such, we assume $P^{\text{ill}}_{\text{generic}} = P^{\text{ill}}_{\text{guaranteed tender}} = P^{\text{ill}}_{\text{Choice}} = P^{\text{ill}}_{\text{CAB}}$. Finally, we did not ask subjects their beliefs about the tenderness of the natural steak, and as such, we assume it equal to the average tenderness belief of the generic, Choice and CAB steaks.

One final consideration in the econometric analysis is the possibility that some unobserved variable might influence both beliefs and choice, resulting in simultaneity bias (Teisl and Roe, 2010). To investigate this possibility, we used the control function approach of Petrin and Train (2010). In particular, we first estimated OLS models where the dependent variables were the beliefs about the safety and tenderness of the steaks and independent variables included measures of gender, age, education, frequency of steak consumption, knowledge of cattle production practices, knowledge of food safety and knowledge of beef quality grades typically purchased. Residuals from these regressions were then entered as right-hand-side variables in the RPL model. When such an approach was employed, we found that the residuals were not statistically significant in the RPL, implying no evidence of belief endogeneity. As such, the analysis that follows only reports results from the RPL.

It should be noted that the approach of Petrin and Train (2010) is similar to that of typical two- or three-stage regressions in the continuous variable setting in that it requires a valid instrument to correctly account for endogeneity. In our context, we utilise perceived knowledge in addition to other variables as instruments. The knowledge variables come from questions which asked 'How knowledgeable do you consider yourself about the following issues? (1 = no knowledge and 7 = very knowledgeable)' and respondents rated themselves on 'cattle production practices', 'USDA Beef Quality Grading System' and 'food safety'. Conceptually, it seems more likely, in a Bayesian-type model, that these variables would relate more to *beliefs* about safety and quality than to the marginal utility of safety/quality. A valid instrument should correlate highly with the variable of interest (beliefs in our case) while being uncorrelated with the residual of the dependent variable. Empirically, the first-stage regressions reveal that the knowledge variables (in addition to some of the demographic variables) are indeed significantly related to beliefs. However, if we enter these variables directly into the utility function during estimation, we find that the knowledge/demographics are *not* significantly related to marginal utilities. Taken together, these results are supportive of the use of instruments employed in this study. Of course, the results might be interpreted with some caution as we cannot claim definitively that there is no endogenity or that the issue is 'solved' given the variables at our disposal, but we hope the exposition here will spur additional research on these topics. One approach that might be considered for supplying exogenous instruments is to randomly give different people different statements about the 'objective' likelihood of outcomes and then elicit subjective beliefs – such an approach was used by Hayes *et al.* (1995) and Teisl and Roe (2010).

3.3. Results and discussion

Table 1 reports summary statistics associated with the subjects' beliefs that the various steaks are tender and will cause illness due to hormones and antibiotics. Overall, subjects believed that the Choice and CAB steaks had a high likelihood of being tender; however, there was substantial variation across respondents, with one subject stating only a 5 per cent chance of the CAB steak being tender and another stating a 98 per cent chance. Subjects also believed the generic steak to be relatively safe to consume, with only a 12 per cent chance, on average, of ultimately causing illness due to growth hormones and antibiotics. Again, there was substantial heterogeneity with some people believing use of hormones and antibiotics were absolutely safe and others believing they were almost certain to cause illness.

Table 2 reports the results of four model specifications, the first three columns report results where only the generic parameter was specified as random in the population and the fourth column reports a specification where all parameters are random. The first column of results presents the standard 'reduced form' RPL results. The second column reports results of Equation (2), which

Table 1. Beliefs about the safety and tenderness of five different steaks elicited from study 1 participants ($n = 37$)

Steak	Mean	SD	Min.	Max.
Probability of tenderness				
Generic	0.466[a]	0.160	0.100	0.750
Guaranteed tender	1.000[b]	–	–	–
Natural	0.665[b]	–	–	–
Choice	0.770[a]	0.150	0.300	1.000
CAB	0.758[a]	0.178	0.050	0.980
Probability of illness				
Generic	0.121[a]	0.186	0.000	0.950
Guaranteed tender	0.121[b]	–	–	–
Natural	0.000[b]	–	–	–
Choice	0.121[b]	–	–	–
CAB	0.121[b]	–	–	–

[a]Values elicited from the post-experiment survey.
[b]Values assumed during model estimation.

Table 2. Results from random RPL models fit to non-hypothetical choice data from study 1

Parameters	Conventional reduced-form RPL	RPL with beliefs	RPL with beliefs and prob. weight	RPL with beliefs and added heterogeneity
−1*Price (mean)	1.546* (0.076)	1.637* (0.082)	1.643* (0.083)	2.699* (0.189)
σ_{price}	–	–	–	0.559* (0.079)
Generic vs. None (mean)	6.690*[a] (0.541)[b]	1.680* (0.817)	1.665* (0.807)	1.855 (1.650)
σ_{generic}	2.329* (0.369)	2.500* (0.416)	2.481* (0.411)	3.813* (0.477)
Guaranteed tender vs. None (mean)	10.813* (0.501)	4.431* (0.946)	4.037* (0.982)	2.845 (2.315)
$\sigma_{\text{guaranteed tender}}$	–	–	–	2.021* (0.351)
Natural vs. None (mean)	10.338* (0.490)	4.343* (0.858)	4.117* (0.853)	3.944* (1.987)
σ_{natural}	–	–	–	3.585* (0.461)
Choice vs. None (mean)	11.519* (0.527)	5.876* (0.856)	5.864* (0.847)	6.228* (1.899)
σ_{Choice}	–	–	–	0.155 (0.405)
CAB vs. none (mean)	11.417* (0.523)	5.79* (0.853)	5.779* (0.844)	5.874* (1.893)
σ_{CAB}	–	–	–	1.035* (0.261)
U(tender) – *U*(tough) (mean)	–	3.000* (0.695)	3.635* (0.866)	8.853* (1.755)
$\sigma_{\text{tender−tough}}$	–	–	–	2.074* (0.788)
U(well) – *U*(ill) (mean)	–	4.655* (0.670)	4.689* (0.669)	10.864* (1.630)
$\sigma_{\text{well−ill}}$	–	–	–	1.552* (0.585)
Probability weighting parm.	–	–	0.770* (0.112)	–
Log-likelihood	−634.6	−595.6	−594.5	−454.1
AIC	1,283.1	1,209.2	1,209.0	940.2
BIC	1,294.4	1,223.7	1,225.1	1,011.3

Number of individuals = 37; number of choices = 629.
[a]Asterisks represent significance at the 0.05 level or lower.
[b]Numbers in parentheses are standard errors.

incorporates beliefs and allows for the identification of preferences for tenderness and healthiness. The third column reports results for Equation (3), which allows for non-linear probability weighting. The final column is the same as column two except that all parameters are random.[4] The AIC and BIC model selection criteria clearly favour the models incorporating beliefs over the reduced form RPL. Comparing models two and three, the AIC selection criteria suggest the best fitting model is the one incorporating non-linear probability weighting, whereas the BIC suggests the opposite. Nevertheless, the estimated value for $\delta = 0.770$ is within the range of estimates commonly found in the literature, and it suggests over-weighting of probabilities <0.41 and under-weighting of probabilities >0.41.

The estimates in Table 2 reveal that accounting for beliefs (either with or without weighting or with all parameters random or not) significantly changes one's interpretation of the results. For example, the conventional 'reduced form' RPL suggests that subjects are WTP an average of USD 2.67 to have a guaranteed tender steak instead of a generic steak [note: 2.67 = (10.813–6.69)/1.546]. In contrast, WTP to have a natural steak instead of generic steak is USD 2.36. Based on these results, one might conclude that consumers more highly value tenderness than hormone-antibiotic safety (note, however, that one cannot reject the null that the mean WTP for guaranteed tender is the same as that for natural). However, results from the models incorporating beliefs show that such a conclusion might be mistaken. The model incorporating beliefs (and assuming linear probability weighting) reveals a WTP for tenderness with a certainty of 3.000/1.637 = USD 1.83 and a WTP for wellness with a certainty of 4.655/1.637 = USD 2.84 [the difference in the two WTP values is significantly different from zero at the $p = 0.04$ level according to the test suggested by Poe, Giraud and Loomis (2005)]. In the model with non-linear probability weighting, these values are USD 2.21 and USD 2.85, respectively (the difference is not statistically significant; $p = 0.17$). The last model with random price, tenderness and wellness parameters suggests the median WTP of USD 3.27 for tenderness and USD 4.02 for wellness.[5] Once beliefs are taken into account, WTP for safety is greater than or equal to WTP for tenderness, depending on the model investigated.

The reason why the conventional 'reduced form' model yields a potentially misleading result is that it does not take into account the fact that most people believe that the generic steak is safe. The reason the premium for natural over generic was so low in the 'reduced form' model was *not* because people did not care about safety but rather because they, on average, *believed* the health

4 We were unable to achieve convergence of a model specified with non-linear probability weighting and all random parameters after a reasonable period of time. As such, we compare specifications where the generic parameter is random, so that one can clearly identify the changes that occur when moving from the conventional reduced form approach to the models with beliefs before presenting the results where all parameters are specified as random.

5 In the final model, WTP is a ratio of two normally distributed random parameters; the ratio itself is not normally distributed. The distribution is not symmetric, so we report the median ratio of 20,000 parameters drawn from the estimated distribution. For reference, the respective mean ratios are USD 3.44 and USD 4.23.

risks from growth hormones and antibiotics in the generic steak were low. One might be tempted to ask 'so what?' If one is only interested in predicting whether consumers will buy a steak labelled 'natural' or 'guaranteed tender' over an unlabelled steak, the reduced form model may be sufficient to provide such insights. However, one misses a great deal of understanding in the reduced form model that might be useful in designing new products or marketing campaigns.

For example, the models incorporating beliefs also allow one to decompose subjects' WTP premiums for natural and guaranteed tender steaks over generic into their constituent parts; something impossible with the reduced form model. Evaluating the model results in the third column in Table 2 at the mean level of beliefs, we find that WTP to have a guaranteed tender steak relative to generic is USD 2.67 (identical to the aforementioned 'reduced form' model estimate of USD 2.67). The advantage of the structural model is that it allows one to determine how much of this premium is due to tenderness and how much is a result of other factors. The estimates show that $(4.037-1.665)/1.645 =$ USD 1.44 of this total is due to factors unrelated to tenderness or safety. Stated differently, of the total WTP premium for guaranteed tender steak, 46 per cent is due to perceived value of added tenderness; the remaining 54 per cent is due to other factors. A similar computation reveals that of the total WTP premium for natural steak over the generic steak, only 38 per cent is due to perceived added healthiness or no hormone use; the remaining 62 per cent is due to other factors.

To further illustrate the insights gained by the structural models, consider a consumer's WTP for the CAB steak. The reduced form model suggests a WTP for CAB (relative to none) of $11.417/1.546 =$ USD 7.38. At the mean level of beliefs, the RPL model with beliefs and probability weighting in the third column indicates a similar overall WTP of $[5.779 + w(0.758) \times 3.635 + w(1-0.121) \times 0.770]/1.643 =$ USD 7.30. With this latter model; however, we can see that of the total USD 7.30 consumers are willing to pay, 48 per cent is due to factors unrelated to tenderness or safety beliefs, 21 per cent is due to perceived value of tenderness and 31 per cent is due to perceived value of wellness. Moreover, the results indicate that if the brand managers of CAB could convince people that CAB was 100 per cent tender, the total value of CAB would increase from USD 7.30 to USD 8.01, a 9.7 per cent increase. Similarly, if CAB went 'all natural' and avoided hormone and antibiotic use, the results indicate that the total value of CAB would increase from USD 7.30 to USD 7.87, a 7.8 per cent increase. These sorts of pragmatic projections would be impossible with the reduced form model.

4. Study 2: effects of beliefs in an experimental auction

The second study seeks to determine the extent to which the result of the first study carry over to an experimental auction environment in which non-hypothetical WTP bids for beef steaks are merged with data collected in a post-experiment survey eliciting beliefs about the safety and tenderness of steaks.

4.1. Procedure

We make use of the experimental auction data collected by Lusk, Feldkamp and Schroeder (2004). In their study, randomly recruited subjects from Manhattan, KS participated in experimental auctions. Subjects were divided into several treatments. About half of the participants were assigned to treatments in which they were given a generic steak and they bid to exchange it with one of the four other steaks previously described in study 1. The other half bid outright to obtain one of the five steaks: generic, guaranteed tender, natural, Choice or CAB. Bids were elicited using incentive-compatible elicitation mechanisms, and subjects participated in a second price auction, a random *n*th price auction, an English auction or the Becker–DeGroot–Marschak (BDM) mechanism. In the data analysis that follows, we pool the data across the four auctions (using the first round of auction bids from the second price and random *n*th price auctions), and we also pool the data across the two endowment frames by focusing on WTP *premiums* to have either the guaranteed tender, natural, Choice or CAB steaks instead of the generic. Hypothesis tests are supportive of pooling the data in this way (see Lusk, 2010; Lusk, Feldkamp and Schroeder, 2004). Exact instructions used in all the treatments can be found in Lusk and Shogren (2007).

At the conclusion of the auctions, subjects completed the exact same survey as in study 1. As described above, the survey elicited subjects' beliefs about the tenderness of the generic, Choice and CAB steaks, and beliefs about the potential health effects of the generic steak resulting from the use of growth hormones and antibiotics. In total, 233 subjects completed the survey and provided usable data for the present analysis. One hundred and sixteen subjects participated in treatments where they were endowed with the generic steak and bid to exchange it with one of the four steaks. In these treatments, the WTP premiums for four steaks were directly elicited, producing a total of 464 bids. Another 117 subjects participated in treatments where they bid outright to obtain each of the five steaks. The WTP premiums for the four non-generic steaks were computed by taking the difference in the bids for those steaks and the bid for the generic steak. In total, there were 468 implied WTP premiums from these treatments, making a total of 932 bids available for analysis.

4.2. Model

To show how subjects' WTP premiums are a function of beliefs, let the expected utility of the generic steak be written as follows:

$$\mathrm{EU}_{i,\mathrm{generic}} = \gamma_{\mathrm{generic}} + (1 - P^{\mathrm{ill}}_{i,\mathrm{generic}})U(\mathrm{well}) + P^{\mathrm{tender}}_{i,\mathrm{generic}}U(\mathrm{tender}) + \alpha y, \quad (4)$$

where y is income and all other variables are previously defined. Now, let the expected utility of a 'higher quality' labelled steak option k be given as follows:

$$\mathrm{EU}_{ik} = \gamma_k + (1 - P_{ik}^{\mathrm{ill}})U(\mathrm{well}) + P_{ik}^{\mathrm{tender}}U(\mathrm{tender}) + \alpha(y - \mathrm{WTP}_{ik}). \quad (5)$$

To find the maximum amount one is WTP to have steak option k rather than the generic steak, set Equation (4) equal to Equation (5) and solve for WTP_{ik}:

$$\mathrm{WTP}_{ik} = \frac{\{(P_{ik}^{\mathrm{ill}} - P_{i,\mathrm{generic}}^{\mathrm{ill}})U(\mathrm{well}) + (P_{ik}^{\mathrm{tender}} - P_{i,\mathrm{generic}}^{\mathrm{tender}})U(\mathrm{tender}) + (\gamma_k - \gamma_{\mathrm{generic}})\}}{\alpha}. \quad (6)$$

Equation (6) shows that the premium one is willing to pay for steak option k relative to the generic steak is a function of the *difference* in beliefs that the respective steaks will cause illness and will be tender, and the difference in the non-wellness, non-tenderness steak characteristics, $\gamma_k - \gamma_{\mathrm{generic}}$. Variation in beliefs across individuals allows for the identification of the parameters $U(\mathrm{well})/\alpha$ and $U(\mathrm{tender})/\alpha$, which reveal WTP for a 100 per cent chance of wellness (relative to illness) and tenderness (relative to toughness), respectively. A conventional 'reduced form' approach cannot separately identify these parameters, as it lumps all the differences in beliefs into the differences in utility: $(\tilde{\gamma}_k - \tilde{\gamma}_{\mathrm{generic}})/\tilde{\alpha}$. In contrast, controlling for beliefs as in Equation (6) allows one to estimate consumers' WTP for wellness, $U(\mathrm{well})/\alpha$ and tenderness, $U(\mathrm{tender})/\alpha$ in addition to the net utility differences.

There are four equations in the system specified by Equation (6) with dependent variables equal to the WTP premiums for the guaranteed tender, natural, Choice and CAB steaks over the generic steak. We use full information maximum likelihood estimation to obtain the model parameters, allowing the error terms across equations to exhibit contemporaneous correlation as a result of the panel nature of the data.[6] As in study 1, we first estimate a standard reduced form model which only provides estimates of $(\tilde{\gamma}_k - \tilde{\gamma}_{\mathrm{generic}})/\tilde{\alpha}$; then we estimate Equation (6), which provides estimates of $(\gamma_k - \gamma_{\mathrm{generic}})/\alpha$, $U(\mathrm{well})/\alpha$ and $U(\mathrm{tender})/\alpha$; and then finally we estimate a specification that allows for probability weighting by transforming the probabilities into weights using a probability weighting function.[7]

6 At first blush, it might appear that some sort of the multivariate Tobit model is the appropriate specification; however, because (for more than half the observations) the dependent variable is the difference in bids, no censoring occurs. In fact, there are multiple negative observations for every dependent variable in our dataset. Analysing the data separately for those treatments where subjects were endowed with a generic steak using Tobit models leads to qualitatively similar results as those discussed in the text.

7 As in study 1, it is possible that simultaneity bias exists if an unobserved variable influences both beliefs and WTP. Using the same instruments described in the control function approach used in study 1; however, the result of a Hausman specification test indicates a failure to reject the null hypothesis that the preferred model is the one without the endogeneity correction.

4.3. Results and discussion

Table 3 reports the results of three model specifications. Both the AIC and BIC model selection criteria reveal that inclusion of beliefs improved the model fit. The final column of the results shows some evidence of non-linear probability weighting, with people over-weighting low-probability events and under-weighting high-probability events; however, inclusion of probability weighting does not improve the AIC or BIC relative to the model with linear beliefs. As a result, we focus our discussion on the middle column of the results when discussing models that incorporate beliefs.

The first column of Table 3 reports conventional reduced form estimates. The parameters represent the mean WTP premiums for each of the four steak types relative to the generic steak. The results indicate that subjects were WTP USD 0.88 to have a guaranteed tender steak relative to the generic steak, and USD 0.55 for the natural steak over the generic steak. As in study 1, this finding would seem to suggest a greater preference on the part of consumers for steak tenderness than avoiding illness from hormones and antibiotics. The second column of results, however, suggests that this conclusion is driven instead by differences in perceptions about the relative safety and tenderness of the generic steak rather than an underlying preference for these attributes, as the values for tenderness and wellness are nearly identical. The middle column of the results shows that consumers are WTP USD 1.28 for a guarantee of tenderness relative to a guarantee of toughness, and are WTP USD 1.28 for a guarantee of no illness from growth hormones or antibiotics.

Interestingly, the parameter estimates associated with the net value of guaranteed tender steak over the generic steak, $(\gamma_{\text{guaranteed tender}} - \gamma_{\text{generic}})/\alpha$, and the net value of the natural steak over the generic steak, $(\gamma_{\text{natural}} - \gamma_{\text{generic}})/\alpha$,

Table 3. Results from seemingly unrelated regression models fit to non-hypothetical WTP bid data from study 2

Parameters	Conventional reduced-form model	Model with beliefs	Model with beliefs and prob. weight
Guaranteed tender vs. Generic	0.882* (0.084)	0.173 (0.221)	−0.107 (0.258)
Natural vs. Generic	0.548* (0.106)	0.043 (0.162)	−0.010 (0.173)
Choice vs. Generic	1.509*[a] (0.115)[b]	1.099* (0.197)	1.099* (0.200)
CAB vs. Generic	2.084* (0.151)	1.630* (0.219)	1.602* (0.219)
$[U(\text{tender}) - U(\text{tough})]/\alpha$	–	1.279* (0.291)	1.710* (0.349)
$[U(\text{well}) - U(\textit{ill})]/\alpha$	–	1.279* (0.535)	1.271* (0.527)
Probability weighting parm.	–	–	0.695* (0.140)
Log-likelihood	−1,587.0	−1,572.9	−1,572.4
AIC	3,182.0	3,157.8	3,158.7
BIC	3,196.7	3,178.5	3,182.9

[a]Asterisks represent statistical significance at the 0.05 level or lower.
[b]Numbers in parentheses are standard errors.
Number of individuals = 233; number of bids = 932.

are not statistically different from zero in the model incorporating beliefs (the estimates are USD 0.17 and USD 0.04, respectively). The implication of these results is that the entire premium for the guaranteed tender and natural steaks over the generic steak is attributable to beliefs/preferences for tenderness and wellness. Other factors (i.e. steak juiciness, environmental impacts etc.) are not apparently a factor in the WTP premium for these steaks according to the data from the experimental auctions.

In addition to providing further understanding behind the factors driving WTP for different steaks, the models incorporating beliefs can also be used to derive practical marketing implications. For example, the reduced form model indicates that consumers are willing to pay a USD 2.08 premium for CAB over generic. At the mean level of beliefs, the results in the second column suggest a nearly identical premium of USD 2.07. The belief data from the auction experiment reveal that people, on average, only believe there is a 45 per cent chance the generic steak is tender, whereas they believe there is an 80 per cent chance the CAB steak is tender. Given that the value of tenderness is USD 1.28, this suggests that of the total USD 2.07 premium $(0.80-0.45) \times 1.28 =$ USD 0.44 (or $\sim$21 per cent) is a result of beliefs about tenderness. If advertising or certification programmes could convince consumers that CAB was 100 per cent tender, the relative premium would increase from USD 2.07 to USD 2.32, a 12.5 per cent increase.

5. Study 3: effects of beliefs in an in-store non-hypothetical choice experiment

The third study moves to a different topic, country of origin meat labelling, and uses data from a non-hypothetical, in-store choice experiment. We focus on how including beliefs in the analysis affects interpretation of consumers' WTP for a meat product with no provenance label relative to one labelled with a definitive origin. As will be illustrated, the statistic that is needed to identify the welfare effects of the labelling policy requires knowledge of consumer beliefs, and once beliefs are taken into account, the apparent value of the origin label falls.

5.1. Procedure

Consumers were recruited from two supermarkets located in suburbs of Dallas, TX and San Antonio, TX during October 2010 and January 2011 as they passed by the fresh meat counter in a grocery store. Subjects were offered a free 12 oz cut of meat (either a beef steak or a pork chop) in addition to a small amount of cash (either USD 2 or USD 4) to participate in the study. We did not find significant differences in choice patterns across the two types of meat studied or the two different levels of cash endowment, and so we pooled the data across these treatments in what follows; for expositional convenience, we also pool the data cross the two locations.

After subjects agreed to participate, they answered nine discrete choice questions. In each of the nine questions, respondents choose from eight options between steaks (or pork chops) from specific origins that differed in terms of cost. The choice was between keeping an unlabelled steak (or pork chop) which respondents had been given for participating in the study or paying a price to instead have one of seven steaks (or pork chops) labelled specifically as being from the USA, Canada, Mexico, Australia (Denmark for pork), Canada and USA, Mexico and USA or Canada, Mexico and USA. Participants were informed that all the meat products were 'USDA inspected and are of the same size, weight, and quality grade' and that the unlabelled steak they had been given for participation 'could be from the U.S., Canada, Mexico, Australia, or a combination of these origins but you will not know exactly which country the steak is from' but that the 'likelihood of a steak being from . . . these origins is similar to the likelihood of finding steaks from one of these locations in a typical grocery store in the U.S.'

Across the nine choices, the prices of the origin-labelled steak options were varied between the values of USD 0, USD 2 and USD 4, whereas the 'keep unlabeled steak' option was the status-quo option equal to a price of USD 0 in each choice (this was the free steak respondents were promised for participating).[8] To make the choice task incentive compatible, a nine-sided die was rolled to determine the binding decision and the choice the participant indicated in the binding scenario determined which steak they received and how much they paid (if anything).

After completing the choice questions, subjects answered a short questionnaire. One question elicited people's beliefs regarding the origin of meat products they encounter in the grocery store.[9] In particular, participants were asked: 'Out of the last 10 beef steaks ≪or pork chops≫ you purchased, how many do you think were from each of the following countries or combination of countries?' Then, we listed each of the origins used in the choice experiment along with a blank for participants to indicate a number. At the bottom of the list of origins, participants were instructed that the sum across all origins should equal exactly 10. We used the number of products a participant said they purchased from a particular origin divided by 10 as a proxy for their belief about the origin of the unlabelled product.

Our data come from 244 subjects, each of whom answered nine discrete choice questions, providing a total sample size of 2,196 choices.

8 Given seven non-status quo steak options, each varying at three price levels, it was necessary to create 27 choices to attain a perfectly orthogonal (i.e. prices are completely uncorrelated with origins) design. Because we thought 27 choices were too many for subjects to complete in an in-store setting, we blocked the questions in three sets of nine. Each subject randomly received one of the blocks.

9 Even though our study took place after mandatory country of origin labelling was in effect, there remained a great deal of uncertainty about the origin of meat products on the market. The labels are typically printed in very small font on the back of the package. Sixty per cent of subjects in this study said they never look for origin labels on meat and ~58 per cent did not know that mandatory origin labelling existed (and another 23 per cent incorrectly said there was no origin labelling law).

5.2. Model

We specify the random utility of labelled steak option j for individual i as follows:

$$\mathrm{EU}_{ij} = U(\mathrm{origin}_j) - \alpha \mathrm{Price}_j + \varepsilon_{ij} \quad \text{for all } j \neq \text{no label}, \tag{7}$$

where $U(\mathrm{origin}_j)$ is the utility derived from origin j (note: for identification purposes, we set the utility of US origin products equal to zero such that the estimated utilities for other origins are in relation to US origin) and α is the marginal utility of income as before. Although Equation (7) is an adequate representation of the utility of a steak option when the origin is precisely known, one of the key questions is consumers' preference for the unlabelled option, for which origin is not precisely known. By incorporating information regarding subjects' beliefs about the origin of an unlabelled product, the random *expected* utility of this option can be specified as follows:

$$\begin{aligned}\mathrm{EU}_{i,\mathrm{unlabelled}} = \theta + \sum_{k=1}^{7} P_{ik} U(\mathrm{origin}_k) - \alpha \mathrm{Price}_{\mathrm{unlabelled}} + e_i \\ + \varepsilon_{i,\mathrm{unlabelled}},\end{aligned} \tag{8}$$

Where θ is an 'uncertainty discount' equal to the (dis)utility of the unlabelled option that cannot be explained by price or by the expected utility of the origin that is given by $\mathrm{EU[origin]}_{\mathrm{unlabelled}} = \sum_{k=1}^{7} P_{ik} U(\mathrm{origin}_k)$, where P_i^k is subject is belief that the unlabelled steak/chop comes from origin k. We add e_i, an individual-specific error term, to capture taste heterogeneity and to account for the panel nature of the data.

Without information on beliefs, a conventional 'reduced form' analysis would specify Equation (8) as $\mathrm{EU}_{i,\mathrm{unlabelled}} = \pi - \alpha \mathrm{Price}_{\mathrm{unlabelled}} + \varepsilon_{i,\mathrm{unlabelled}}$, where the parameter π would capture the *joint* effect of the uncertainty discount and the expected utility of origin. Aside from the potential problem that π may be a biased estimate of $\theta + \sum_{k=1}^{7} P_{ik} U(\mathrm{origin}_k)$ due to the omission of individual-specific beliefs that may be correlated with the error term, knowledge of this single coefficient does not permit one to decompose consumers' WTP to avoid an unlabelled option into the constituent parts of θ and $\sum_{k=1}^{7} P_{ik} U(\mathrm{origin}_k)$. Mandatory labelling policies, however, will only eliminate the disutility associated with uncertainty, θ, and this value cannot be cleanly derived unless one has knowledge of beliefs. Stated differently, a conventional reduced form analysis cannot yield an estimate of the statistic needed to help assess the value of country of origin labelling.

Given Equations (7) and (8) and the assumption that the ε_{ij} are distributed iid extreme value type I and e_i is normally distributed, an RPL can be estimated. As in the previous two studies, we also estimate a specification allowing for non-linear probability weighting. However, unlike the models in studies 1 and 2 in which there were only two outcomes (i.e. tender vs. tough), there is a possibility that the unlabelled steak could take on one of seven possible origins. The rank-dependent model requires that outcomes be ordered in terms of their relative desirability with cumulative probabilities entering the probability weighting function (see Quiggin, 1982). To determine the preference order, we first estimated the conventional reduced form model and used those estimates to rank the origins in terms of their relative desirability. Let $k = 1$ indicate the most desirable location, $k = 2$ indicate the second most desirable location and so on. Now, the random rank-dependent expected utility of the unlabelled option can be written as follows:

$$\mathrm{RDEU}_{i,\text{unlabelled}} = \theta + \sum_{k=1}^{7} z_{ik} U(\text{origin}_k) - \alpha \text{Price}_{\text{unlabelled}} + e_i + \varepsilon_{i,\text{unlabelled}}, \tag{9}$$

where $z_{ik} = w\left(\sum_{t=1}^{k} P_{it}\right) - w\left(\sum_{t=1}^{k-1} P_{it}\right)$ and where w is the weighting function previously described. When $k = 1$, $z_{i1} = w(P_{i1})$, and when $k = 7$, $z_{i7} = 1 - w\left(\sum_{t=1}^{6} P_{it}\right)$.

Finally, as in the previous two studies, we considered the possibility of simultaneity bias resulting from some unobserved variable that might influence both beliefs and choice, using the Petrin and Train (2010) control function approach. We first estimated OLS models where the dependent variables were the beliefs about the likelihood of a steak coming from different origins and independent variables included measures of gender, age, education, frequency of beef/pork consumption, knowledge of country of origin labelling law and race. Residuals from these regressions were then entered as right-hand-side variables in Equation (8). Three of the residuals were statistically significant at the 0.05 level, and as such, we report the 'corrected' results along with the other estimates in what follows. As before, we also present results allowing for different degrees of heterogeneity in the estimated preference parameters.

5.3. Results and discussion

The data from the questionnaire reveal that the subjects believe that, on average, 73.8 per cent of the beef/pork they buy comes from the USA. Given that our instructions told participants that the exact origin they would receive if they picked the unlabelled option was proportional to the origins sold in grocery stores, we interpret these data to imply subjects' believe there is a 0.738 probability of the unlabelled option in the experiment being from US origin. The next most likely category was Mexico with a mean probability of 0.07 followed by USA, Canada and Mexico at 0.064. Subjects believed it least likely that meat came from Australia in the case of beef (Denmark in the case of pork), stating a mean probability of 0.016, and from USA and Mexico with a mean probability equal to 0.013.

Table 4 reports the results for four model specifications in which the disutility of the unlabelled origin option is specified as random in the population. All four models reveal that consumers most prefer US origin meat and least prefer Mexican origin meat. According to the AIC measures of fit, only the RPL model accounting for endogenous beliefs better fits the data than the conventional reduced form RPL. Yet, the BIC suggests that the reduced form model is the best fitting model overall. However, as previously argued, there is reason to suspect that the composite parameter $\pi = -4.695$ in the reduced form model is biased as it is likely correlated with unobserved variables (i.e. beliefs) that are also correlated with the error term. In such cases, goodness of fit is not necessarily a reliable indicator of which model properly identifies the true structural parameters. The reduced form model reveals that consumers are WTP a total of USD 9.01 to have a US vs. an unlabelled steak, a value that is somewhat lower than that implied by the preferred model with beliefs, which range from USD 9.30 to USD 10.75.

The advantage of the models incorporating beliefs is that they allow one to decompose the composite disutility of the unlabelled steak and identify the value of the disutility associated with uncertainty in the unlabelled option. In particular, results indicate that of the total amount consumers are WTP for the unlabelled steak, USD 7.61 (note: USD 7.61 = 3.969/0.521) or 82 per cent of the total discount for the unlabelled option is due to the disutility associated with having uncertainty regarding the origin label. While these values are non-trivial, they represent a lower value prospect than what is implied by the composite WTP value to avoid the unlabelled product relative to the US product. As indicated, a conventional reduced form approach might assert that the value of labelling is USD 9.01; however, the structural model with beliefs shows resolving the disutility from uncertainty is only worth USD 7.61, an amount that is USD 1.39 (or 15.4 per cent) lower than that might be implied from a reduced form analysis. Moreover, the results reveal that, at the mean beliefs, consumers are WTP a premium of only about USD 1.68 (note: USD 1.68 = 0.877/0.521) for a US origin steak relative to the 'weighted average origin' steak. The reason why the value is so low is that most people believe the unlabelled steak is highly likely to come from US origin.

Finally, given the expected utility framework, it is possible to utilise results in Table 4 to draw insights that would be missed in conventional analysis. In particular, three of the origin options involve combinations of the other origins. As such, it is possible to 'back out' subjects' beliefs about the extent to which a combined option is coming from each origin. To illustrate, the last column of results show that people's utility for Canadian origin is -2.863 and US and Canadian origin is -2.074, and recall that utility for US origin has been normalised to zero. The utility of a combined US and Canada label can be conceptualised as arising from the expected utility of its components: $U(\text{USA and CANADA}) = U(\text{USA})p^{\text{USA}} + U(\text{Canada})(1 - p^{\text{USA}})$, where p^{US} is the probability people believe the jointly labelled product comes from the USA. Plugging in the coefficients, the following equality is implied: $-2.074 = 0p^{\text{USA}} - 2.863(1 - p^{\text{USA}})$,

Table 4. Results from PRL models fit to non-hypothetical choice data from study 3

Parameters	Conventional reduced-form RPL	RPL with beliefs	RPL with beliefs and prob. weight	RPL with endogenous beliefs[a]
$EU[origin]_{unlabelled} + \theta^{b}$ vs. USA[c]	−4.695*[d] (0.392)[e]	−4.951* (0.425)	−5.613* (0.643)	−4.846* (0.415)
$EU[origin]_{unlabelled}$ vs. USA	–	−0.866* (0.024)	−1.011* (0.513)	−0.877* (0.025)
θ^{b}	–	−4.085* (0.423)	−4.610* (0.608)	−3.969* (0.413)
St. D $[EU[origin]_{unlabelled} + \theta$ vs. USA]	4.171* (0.353)	4.527* (0.364)	5.318* (0.643)	4.087* (0.360)
Canada vs. USA	−2.862* (0.112)	−2.862* (0.112)	−2.865* (0.112)	−2.863* (0.112)
Mexico vs. USA	−4.883* (0.273)	−4.748* (0.255)	−4.775* (0.258)	−4.888* (0.273)
Australia/Denmark vs. USA	−2.921* (0.114)	−2.919* (0.114)	−2.923* (0.114)	−2.921* (0.114)
USA and Canada vs. USA	−2.074* (0.084)	−2.074* (0.084)	−2.076* (0.084)	−2.074* (0.084)
USA and Mexico vs. USA	−3.696* (0.156)	−3.696* (0.156)	−3.670* (0.156)	−3.696* (0.156)
USA, Canada and Mexico vs. USA	−3.166* (0.126)	−3.149* (0.125)	−3.154* (0.125)	−3.166* (0.126)
−1*Price	0.521* (0.024)	0.520* (0.024)	0.522* (0.024)	0.521* (0.024)
Probability weighting parm.	–	–	0.883* (0.333)	–
Log-likelihood	−2,330.7	−2,336.8	−2,334.6	−2,322.0
AIC	4,676.4	4,691.5	4,689.3	4,674.0
BIC	4,710.9	4,723.0	4,724.2	4,726.5

[a]This model included seven additional variables corresponding to the residuals from seven regressions of demographic and knowledge variables regressed on beliefs about the steak coming from USA, Canadian, Mexican, Australian or Mixed origins. Three of these variables were significant at the 0.05 level, and a likelihood ratio test indicates that the model with endogenous beliefs is preferred to the one without.
[b]The parameter θ corresponds to the disutility associated with uncertainty in the origin of the unlabelled option.
[c]In the conventional model, this parameter is directly estimated; in the models with beliefs, the estimates are calculated at the mean levels of beliefs and the standard errors are calculated via the delta method.
[d]Asterisks represent statistical significance at the 0.05 level or lower.
[e]Numbers in parentheses are standard errors; number of individuals = 244; number of choices = 2,196.

and solving for the probability yields $p^{\mathrm{USA}} = 0.27$. Thus, when a product has a joint US and Canadian origin label, people implicitly believe it to primarily come from Canada (with probability $1-0.27 = 0.73$). Applying the same logic to the US and Mexican origin label reveals people overwhelmingly believe meat from this combined label to come from Mexico, with a probability of 0.76.

The implication is that when a product has a mixed-origin label, people are apparently pessimistic, believing the joint-labelled product to have a much higher likelihood of coming from the less-preferred origin. Retailers wishing to counteract this pessimism might inform consumers of the true probabilities (or processes by which products come to possess joint-origin labels). That consumers' view mixed origins pessimistically may partially explain why retailers refrained from voluntarily providing mixed-origin labels prior to passage of the mandatory country of origin labelling law, and why now many print such labels in small font on the back of packages. If consumers (perhaps incorrectly) perceive mixed origins as signals of likely sourcing by the 'inferior' country listed, then the simple provision of a mixed label may reduce sales more than anticipated using 'objective' probabilities.

Table 5 reports results of a model specification incorporating beliefs, and where every estimated parameter is assumed normally distributed in the population. The results reveal qualitatively similar results to those previously presented in that 18.7 per cent (1.41/7.533) of the total WTP to avoid the unlabelled steak can be explained by beliefs about the expected origin of the steak, whereas 81.3 per cent (6.122/7.533) is a result of the aversion to uncertainty of not knowing the origin. Likewise, when we 'back out' subjects' beliefs about the extent to which a combined Mexican, US option is coming

Table 5. Results from PRL model with beliefs and additional heterogeneity from study 3

Parameters	Mean	SD
$\mathrm{EU[origin]}_{\mathrm{unlabelled}} + \theta^{a}$ vs. USA[b]	−7.533*[c] (0.646)[d]	–
$\mathrm{EU[origin]}_{\mathrm{unlabelled}}$ vs. USA	−1.410* (0.040)	–
θ^{a}	−6.122* (0.491)	6.007* (0.486)
Canada vs. USA	−4.483* (0.294)	2.018* (0.317)
Mexico vs. USA	−7.011* (0.735)	1.781* (0.620)
Australia/Denmark vs. USA	−6.380* (0.504)	4.084* (0.398)
USA and Canada vs. USA	−3.829* (0.277)	2.725* (0.239)
USA and Mexico vs. USA	−5.276* (0.315)	1.033* (0.353)
USA, Canada and Mexico vs. USA	−5.396* (0.359)	2.834* (0.317)
−1*Price	1.053* (0.077)	0.886* (0.082)
Log-Likelihood	−1836.3	
AIC	3704.5	
BIC	3795.7	

Number of individuals = 244; number of choices = 2,196.
[a]The parameter θ corresponds to the disutility associated with uncertainty in the origin of the unlabelled option.
[b]The estimates are calculated at the mean levels of beliefs and the standard errors are calculated via the delta method.
[c]Asterisks represent statistical significance at the 0.05 level or lower.
[d]Numbers in parentheses are standard errors.

from each origin, we find that people believe the steak to come from Mexico with a probability of 0.75, which is almost exactly what was found in Table 3.

6. Conclusions

This paper reported three studies in which consumers' beliefs were directly incorporated into the econometric estimation to yield structural models of consumer preferences. In all three studies, we showed that incorporating knowledge of beliefs improves the understanding of consumer behaviour and yields important insights that would be unavailable in the absence of belief information.

Studies 1 and 2 both revealed that without considering consumers' beliefs, we are tempted to mistakenly conclude that one attribute (tenderness) is more important than another (avoidance of illness from growth hormones and antibiotics). Moreover, the analysis revealed information about the potential value of changing consumers' beliefs about different quality characteristics. The result from study 1 suggested that a portion of people's WTP for guaranteed tender and 'natural' steaks can be explained by factors other than tenderness and concerns over growth hormones, although study 2 suggested this might represent a small amount. Both studies provided some evidence of non-linear probability weighting with people over-weighting low-probability events and under-weighting medium to high-probability events. Studies 1 and 2 yielded qualitatively similar insights insofar as the implications about the role of beliefs. These similarities were observed despite very different frames in which consumer preferences were elicited in the two studies (i.e. choice experiment vs. experimental auction). The third study revealed that the data on beliefs were necessary to isolate the disutility consumers experience when purchasing steaks and pork chops without any origin information, and the results suggest that attempts to calculate consumer WTP to avoid unlabelled steaks are likely to exaggerate the value of a labelling policy.

The take-home message of this study is that experiments and surveys that are designed for *preference elicitation* need to be combined with the often-ignored process of *belief elicitation*. If one has data only on choice or WTP, it impossible to isolate preferences from beliefs. However, given data on choice (or WTP) and beliefs, one can then back out information on the third construct: preferences. An alternative approach, shown in study 3 and taken by Lusk (2011), is to use a survey approach to measure preferences, and then use choice data to back out implied beliefs. Economists have become increasingly creative at developing methods to elicit consumer choices and WTP in a way that avoids behavioural biases. It is time to deepen our understanding of consumer choice for food and environmental products by applying that same ingenuity to finding robust ways of measuring consumers' beliefs (see the papers introduced by Bellemare and Manski, 2011).

Eliciting belief information is important. As revealed in our first two studies, there are attributes that people strongly prefer (i.e. safety) but might be perceived as ubiquitous in the market place. When this is true, studies could reveal a low WTP for the attribute even though consumers generally value the trait. From a public policy perspective, such a result might imply significant under-valuation of important attributes based solely on WTP studies. In such cases, even small degradations in trust or beliefs about safety might result in 'unexpectedly' large changes in purchase patterns.

Our studies used hypothetical questions to elicit beliefs, but future work might consider the effects of using scoring rules on elicited beliefs. Moreover, existing research reveals that people often have difficulty making probability judgements, and various elicitation devices or training procedures might lead to more reliable belief estimates (Corso, Hammitt and Graham, 2001). One possibility is to utilise advances in 'best worst' scaling, which uses discrete choice questions to place items on a ratio scale, to determine the relative likelihood of occurrence of different events (e.g. Lusk and Briggeman, 2009; Erdem, Rigby and Wossink, 2012). Our study also uncovered evidence of non-linear probability weighting, but more robust estimates of this phenomenon would likely require repeated decisions (or belief elicitations) over a wide range of possible outcomes, from very unlikely to highly likely (e.g. Abdellaoui, 2000; Drichoutis and Lusk, 2012). Another interesting area for future research relates to the effects of information on beliefs, preferences or both. The standard assumption is that information alters beliefs (e.g. Hayes *et al.*, 1995), but whether this is true and whether people incorporate new information in a Bayesian manner remain open questions (Holt and Smith, 2009).

While the applications studied in this paper relate specifically to meat, the implications extend to other food issues. Given that the vast majority of consumers now have little direct involvement in food production, one might suspect that many food choices are made with inaccurate beliefs regarding production claims. For examples, Norwood and Lusk (2011) report that, on average, consumers believe only 37 per cent of eggs are produced in the cage system when the reality is closer to 95 per cent, and Chang and Lusk (2009) report that, on average, consumers believe small farmers derive larger benefits from transitioning to organic production than do larger farmers. Indeed, many in the agricultural production community bemoan the public's lack of understanding of commercial agriculture and argue for more 'science based' regulation. Although agricultural producer groups differ widely from the average food consumer in their beliefs about the safety and quality of GM foods, irradiation technology and antibiotic use in animal agriculture, just to give a few examples, it is clear that an understanding of these controversies, not to mention the impacts of public policies, requires a better understanding of producer and consumer beliefs.

References

Abdellaoui, M. (2000). Parameter-free elicitation of utility and probability weighting functions. *Management Science* 26: 1497–1512.

Adamowicz, W. L. (2004). What's it worth? An examination of historical trends and future directions in environmental valuation. *Australian Journal of Agricultural and Resource Economics* 48: 419–443.

Bellemare, C. and Manski, C. F. (2011). Introduction: 'measurement and analysis of subjective expectations'. *Journal of Applied Econometrics* 26: 351.

Bellemare, M. F. (2012). Insecure land rights and share tenancy: evidence from Madagascar. *Land Economics* 88: 155–180.

Birnbaum, M. H. and Chavez, A. (1997). Tests of theories of decision making: violations of branch independence and distribution independence. *Organizational Behavior and Human Decision Processes* 71: 161–194.

Buchanan, J. M. (1991). *The Economics and the Ethics of Constitutional Order*. Ann Arbor: University of Michigan Press.

Camerer, C. F. and Ho, T.-H. (1994). Violations of the betweenness axiom and nonlinearity in probability. *Journal of Risk and Uncertainty* 8: 167–196.

Cameron, T. A. (2005). Individual option prices for climate change mitigation. *Journal of Public Economics* 89: 283–301.

Cameron, T. A., DeShazo, J. R. and Johnson, E. H. (2011). Scenario adjustment in stated preference research. *Journal of Choice Modeling* 4: 9–43.

Chang, J. B. and Lusk, J. L. (2009). Fairness and food choice. *Food Policy* 34: 483–491.

Corso, P. S., Hammitt, J. K. and Graham, J. D. (2001). Valuing mortality-risk reduction: using visiual aids to improve the validity of continent valuation. *Journal of Risk and Uncertainty* 23: 165–184.

Dannenberg, A. (2009). The dispersion and development of consumer preferences for genetically modified food: a meta-analysis. *Ecological Economics* 68: 2182–2192.

Dannenberg, A., Scatasta, S. and Strum, B. (2011). Mandatory versus voluntary labelling of genetically modified food: evidence from an economic experiment. *Agricultural Economics* 42: 373–386.

Delavande, A., Ginéc, X. and McKenziec, D. (2011). Measuring subjective expectations in developing countries: a critical review and new evidence. *Journal of Development Economics* 94: 151–163.

Drichoutis, A. and Lusk, J. L. (2012). What Can Multiple Price Lists Really Tell Us About Risk Preferences? MPRA Paper No. 42128, posted 22 October, 2012.

Erdem, S., Rigby, D. and Wossink, A. (2012). Using best-worst scaling to explore perceptions of relative responsibility for ensuring food safety. *Food Policy* 37: 661–670.

Fishbein, M. and Ajzen, I. (1975). *Belief, Attitude, Intention and Behavior*. Reading, MA: Addison-Wesley.

Grunert, K. G. (2005). Food quality and safety: consumer perception and demand. *European Review of Agricultural Economics* 32: 369–391.

Grunert, K. G., Juhl, H. J., Esbjerg, L., Jensen, B. B., Bech-Larsen, T., Brunsø, K. and Madsen, C. Ø. (2009). Comparing methods for measuring consumer willingness to Pay for a basic and an improved ready made soup product. *Quality and Preference* 20: 607–619.

Hayes, D. J., Shogren, J. F., Shin, S. U. and Kliebenstein, J. B. (1995). Valuing food safety in experimental auction markets. *American Journal of Agricultural Economics* 77: 40–53.

Hey, J. D. and Orme, C. (1994). Investigating generalizations of expected utility theory using experimental data. *Econometrica* 62: 1291–1326.

Holt, C. A. (1986). Scoring-rule procedures for eliciting subjective probability and utility functions. In P. K. Goel and A. Zellner (eds), *Bayesian Inference and Decision Techniques –Essays in Honor of Bruno de Finetti*. Amsterdam: Elsevier Science Publishers.

Holt, C. A. and Smith, A. M. (2009). An update on Bayesian updating. *Journal of Economic Behavior and Organization* 29: 125–134.

Hora, S. C., Hora, J. A. and Dodd, N. G. (1992). Assessment of probability distributions for continuous random variables: a comparison of the bisection and fixed value methods. *Organizational Behavior and Human Decision Processes* 51: 133–155.

Huffman, W. E., Rousu, M., Shogren, J. F. and Tegene, A. (2007). The effects of prior beliefs and learning on consumers' acceptance of genetically modified foods. *Journal of Economic Behavior and Organization* 63: 193–206.

Kahneman, D. and Sugden, R. (2005). Experienced utility as a standard of policy evaluation. *Environmental and Resource Economics* 32: 161–181.

Kahneman, D. and Tversky, A. (eds) (2000). *Choices, Values, and Frames*. Cambridge, UK: Cambridge University Press.

Karni, E. (2009). A mechanism for eliciting probabilities. *Econometrica* 77: 603–606.

Lagerkvist, C. J. and Hess, S. (2011). A meta-analysis of consumer willingness to pay for farm animal welfare. *European Review of Agricultural Economics* 38: 55–78.

Leggett, G. G. (2002). Environmental valuation with imperfect information: the case of the random utility model. *Environmental and Resource Economics* 23: 343–355.

Loureiro, M. L. and Umberger, W. J. (2007). A choice experiment model for beef: what US consumer responses tell us about relative preferences for food safety, country-of-origin labeling and traceability. *Food Policy* 32: 496–514.

Louviere, J. J., Hensher, D. A. and Swait, J. D. (2000). *Stated Choice Methods: Analysis and Application*. Cambridge: Cambridge University Press.

Lusk, J. L. (2010). Experimental auction markets for studying consumer preferences. In H. MacFie and S. Jaeger (eds), *Consumer Driven Innovation in Food and Personal Products*. Cambridge, UK: Woodhead.

Lusk, J. L. (2011). External validity of the food values scale. *Food Quality and Preference* 22: 452–462.

Lusk, J. L. and Briggeman, B. C. (2009). Food values. *American Journal of Agricultural Economics* 91: 184–196.

Lusk, J. L. and Coble, K. O. (2005). Risk perceptions, risk preference, and acceptance of risky food. *American Journal of Agricultural Economics* 87: 393–405.

Lusk, J. L., Feldkamp, T. and Schroeder, T. C. (2004). Experimental auction procedure: impact on valuation of quality differentiated goods. *American Journal of Agricultural Economics* 86: 389–405.

Lusk, J. L., House, L. O., Valli, C., Jaeger, S. R., Moore, M., Morrow, B. and Traill, W. B. (2004). Effect of information about benefits of biotechnology on consumer acceptance of genetically modified food: evidence from experimental auctions in United States, England, and France. *European Review of Agricultural Economics* 31: 179–204.

Lusk, J. L. and Schroeder, T. C. (2004). Are choice experiments incentive compatible? A test with quality differentiated beef steaks. *American Journal of Agricultural Economics* 86: 467–482.

Lusk, J. L. and Shogren, J. (2007). *Experimental Auctions: Methods and Applications in Economic and Marketing Research*. Cambridge, UK: Cambridge University Press.

Manski, C. F. (2004). Measuring expectations. *Econometrica* 72: 1329–1376.

Marette, S., Roe, B. E. and Teisl, M. (2012). The welfare impact of food pathogen vaccines. *Food Policy* 3: 86–93.

Norwood, F. B. and Lusk, J. L. (2011). *Compassion by the Pound: The Economics of Farm Animal Welfare*. Oxford, UK: Oxford University Press.

O'Hagan, A., Buck, C. E., Daneshkhah, A., Eiser, J. R., Garthwaite, P. H., Jenkinson, D. J., Oakley, J. E. and Rakow, T. (2006). *Uncertain Judgements: Eliciting Experts' Probabilities*. West Sussex, England: Wiley.

Pennings, J. M. E. and Wansink, B. (2004). Channel contract behavior: the role of risk attitudes, risk perceptions, and channel members' market structures. *Journal of Business* 77: 697–723.

Pennings, J. M. E., Wansink, B. and Meulenberg, M.T.G. (2002). A note on modeling consumer reactions to a crisis: the case of the mad cow disease. *International Journal of Research in Marketing* 19: 91–100.

Petrin, A. and Train, K. (2010). A control function approach to endogeneity in consumer choice models. *Journal of Marketing Research* 47: 3–13.

Poe, G. L., Giraud, K. L. and Loomis, J. B. (2005). Computational methods for measuring the difference of empirical distributions. *American Journal of Agricultural Economics* 87: 353–365.

Prelec, D. (2004). A bayesian truth serum for subjective data. *Science* 306: 462–466.

Quiggin, J. (1982). A theory of anticipated utility. *Journal of Economic Behavior and Organization* 3: 323–343.

Roberts, D., Boyer, T. and Lusk, J. L. (2008). Preferences for environmental quality under uncertainty. *Ecological Economics* 66: 584–593.

Savage, L. J. (1954). *The Foundations of Statistics*. New York: Wiley.

Savage, L. J. (1971). Elicitation of personal probabilities and expectations. *Journal of the American Statistical Association* 66: 783–801.

Schroeder, T. C., Tonsor, G. T., Pennings, J. M. E. and Mintert, J. (2007). Consumer food safety risk perceptions and attitudes: impacts on beef consumption across countries. *B.E. Journal of Economic Analysis & Policy* (Contributions), 7: Article 65.

Shleifer, A. (2012). Psychologists at the gate: a review of Daniel Kahneman's thinking, fast and slow. *Journal of Economic Literature* 50: 1080–1091.

Spetzler, C. S. and von Holstein, C. A. S. S. (1975). Probability encoding in decision analysis. *Management Science* 22: 340–352.

Teisl, M. F. and Roe, B. E. (2010). Consumer willingness-to-pay to reduce the probability of retail foodborne pathogen contamination. *Food Policy* 35: 521–530.

Tversky, A. and Kahneman, D. (1992). Advances in prospect theory: cumulative representation of uncertainty. *Journal of Risk and Uncertainty* 5: 297–323.

Viscusi, W. K. and Hakes, J. (2003). Risk ratings that do not measure probabilities. *Journal of Risk Research* 6: 23–43.

von Neumann, J. and Morgenstern, O. (1944). *Theory Games and Economic Behavior*. Princeton, NJ: Princeton University Press.

Wu, G. and Gonzalez, R. (1996). Curvature of the probability weighting function. *Management Science* 42: 1676–1690.