Measuring the consumer benefits of improving farm animal welfare to inform welfare labelling

Article

Accepted Version

It is advisable to refer to the publisher's version if you intend to cite from the work. See Guidance on citing.

www.reading.ac.uk/centaur

CentAUR

Measuring the consumer benefits of improving farm animal welfare to inform welfare labelling

A. Kehlbacher[a,*], R. Bennett[a], K. Balcombe[a]

[a]*Department of Food Economics and Marketing, School of Agriculture Policy and Development, University of Reading, Earley Gate, PO Box 237, Reading, RG6 6AR, England*

Abstract

Policy makers in the European Union are envisioning the introduction of a community farm animal welfare label which would allow consumers to align their consumption habits with their farm animal welfare preferences. For welfare labelling to be viable the market for livestock products produced to higher welfare standards has to be sufficiently segmented with consumers having sufficiently distinct and behaviourally consistent preferences. The present study investigates consumers' preferences for meat produced to different welfare standards using a hypothetical welfare score. Data is obtained from a contingent valuation study carried out in Britain. The ordered probit model was estimated using Bayesian inference to obtain mean willingness to pay. We find decreasing marginal WTP as animal welfare levels increase and that people's preferences for different levels of farm animal welfare are sufficiently differentiated making the introduction of a labelling scheme in the form of a certified rating system appear feasible.

Keywords: farm animal welfare, labelling, welfare score, meat, contingent valuation, Bayesian

1. Introduction

Farm animal welfare is an emotive issue surrounding the consumption of meat and its protection is important to European citizens (European Commission, 2007). Accordingly, demand for welfare-friendly products has increased in recent years. The British retailer Sainsbury's, for example, reported in 2010 a 164% increase in year-on-year sales of products produced under the Freedom Food Scheme, a farm assurance and food labelling scheme encouraging farmers to adopt and maintain higher welfare standards (RSPCA, 2010). Moreover, 41% of respondents to a survey commissioned by the European Commission (2007) replied 'yes probably' or 'yes definitely' when asked whether they wanted to be better informed about the conditions under which animals are farmed; and the vast majority of British consumers said that they would use product labels as a primary source of information if it were made available to them in that way (Mayfield et al., 2007). While some European countries have voluntary welfare labelling schemes such as Neuland in Germany, Label Rouge in France, or Freedom Food in the UK, no comprehensive European animal welfare labelling scheme exists. Policy makers have addressed this lack of a transparent and credible information system that would enable consumers to act on their animal welfare concerns by adopting the Community Action Plan on the Protection and Welfare of Animals 2006-2010 which envisions, amongst other things, the setting up of a community farm animal welfare label.

A welfare label can take on different formats including that of a certificate certifying the achievement of certain minimum welfare standards and/or a rating system indicating the welfare level to which products have been produced. For both labelling formats to be viable and desirable, the market for meat products produced to higher welfare standards has to be sufficiently segmented with consumers having sufficiently distinct and behaviourally consistent preferences. This paper investigates this issue by looking at British consumers' preferences for meat produced to different levels of animal welfare. A contingent valuation study was carried out using the multiple-bounded uncertainty choice format (Welsh and Poe, 1998) to elicit British meat consumers' willingness to pay (WTP) for meat produced to different animal welfare levels. Unlike previous studies that valued farm animal welfare in terms of changes in husbandry practices, this study uses a welfare score to depict different animal welfare levels, an approach that allows generic value estimates for meat produced to different farm animal welfare standards to be obtained. The score is based on the Welfare Quality® Index which is a recently developed farm animal welfare assessment scheme that is to form the basis of a European standard for the evaluation of animal welfare on farms, during transport and at slaughter (Botreau et al., 2007a).

If a farm achieves a high Welfare Quality® class its animals are free from experiencing periods of prolonged hunger or thirst, thermal discomfort, restricted movement, injuries, disease, pain or fear; and are allowed to express natural, non-harmful, behaviours. This study obtains WTP for animal friendly meat as measured by this index which is likely to form the basis of a standard tool for welfare labelling in the European Union (EU). We explore how people would value changes in the level of a welfare index in monetary terms and how the results can be used to inform policies in particular welfare labelling. In doing so we discuss the potential problems associated with the use of scores in contingent valuation studies.

We conclude that given the level of differentiation of people's preferences for different levels of farm animal welfare, the introduction of a labelling scheme in the form of a certificate or a rating system appears generally feasible as respondents are willing to pay more for meat with higher welfare scores. The derived WTP estimates are affected by respondents' perceptions of the current farm animal welfare situation in the UK with WTP for very high welfare scores being relatively smaller suggesting that welfare labels indicating welfare standards that exceed consumers' satiation levels for farm animal welfare may not be able to command a sufficiently high price premium.

2. Background to this research

2.1. Welfare labelling

Farm animal welfare is a credence good related to the production, transport and slaughter of farm animals that cannot be discerned through inspection of the good before purchase, or through use after purchase (Darby and Karni, 1973). While livestock producers know about their animals' welfare state, consumers are, under usual circumstances, not able to observe or verify it. In the absence of a credible and reliable welfare labelling scheme, consumers are uncertain about welfare levels in livestock production. Livestock producers providing high animal welfare standards therefore often cannot achieve a price premium and thus have no incentive to continue their practice. Hence, information asymmetry in the livestock market leads to high welfare products being underprovided and thus to market failure (Akerlof, 1970). By establishing a welfare labelling scheme that is accredited by a trusted body and has its certification and auditing process monitored, the credence attribute animal welfare can be transformed into a search attribute thus eliminating the information asymmetry and creating a market for animal welfare (Lusk et al., 2007).

A welfare label can take the form of a logo certifying whether certain minimum requirements with regard to animal welfare have been fulfilled by a farm. This labelling format is preferred by 35% of Eurobarometer respondents (European Commission, 2007). A tiered welfare label or score indicating relative animal welfare levels on livestock products is the preferred labelling format for 26% of Eurobarometer respondents (European Commission, 2007). While a certificate requires farms to achieve a given level of farm animal welfare, a tiered system facilitates livestock producers' entrance because they can start at a lower tier. A score thus has the advantage of being an equitable way of including all farms into the welfare labelling scheme. In addition, a tiered system which is reflected in the prices obtained by farmers, provides an incentive to farmers to improve to the next higher tier (Deimel et al., 2010). It would therefore produce additional benefits in terms of improved animal welfare obtained from a given policy measure (Lusk et al., 2007). Furthermore, consumers can buy products according to their preferred level of animal welfare which means that greater product differentiation is possible in comparison to a logo and differing preferences for animal welfare standards can be absorbed. The drawbacks of a welfare score include that it may be more demanding for consumers to comprehend; the communication of differences in welfare levels may be difficult; and welfare levels just above the legal minimum may have little credibility (Deimel et al., 2010).

A welfare labelling scheme can be either mandatory or voluntary. If only a small segment of the population is interested in knowing about the welfare status of the animals from which their livestock products are derived and are willing to pay more for them, a voluntary labelling scheme is more efficient. In this instance, individual producers or producer groups would be left to decide whether they use the welfare labelling scheme and label their products accordingly (FAWC, 2006). This means that the welfare labelling scheme's standard can be set at a level that largely surpasses the welfare standards customary within the industry. However, a voluntary labelling scheme in the form of a score entails the risk that the services of farmers in the lowest tier may not be rewarded as consumers equate their efforts to improve animal welfare with the legal minimum standards (Deimel et al., 2010). By contrast, a mandatory labelling scheme is more effective if the majority of the public is interested in knowing the welfare status of their livestock products. However, because the entire market has to be segregated and labelled even though only a portion of consumers may care about the attribute in question, mandatory labelling comes at a higher cost than voluntary labelling (Caswell, 1998). If retailers were required to indicate the animals' welfare state on all livestock products, it is possible that in the long term consumer preferences would shift to higher welfare standards which in turn would feed back to livestock producers and potentially benefit the animals (FAWC,

2006).

To date, the authors have found only one study which has attempted to value the benefits of a farm animal welfare label. Nocella et al. (2010) carried out an internet survey in five European Union (EU) countries asking consumers for their WTP to buy certified animal friendly products. The present study investigates British consumers' preferences for meat labelled with differing welfare scores. Our results can help inform policy makers about where to set the level of a minimum animal welfare standard certified by a certificate that maximises consumer benefits. In addition, knowing consumers' WTP for different welfare levels can inform policy makers who need to establish whether consumer preferences for different levels of farm animal welfare are sufficiently differentiated so as to warrant the introduction of a tiered welfare labelling scheme or a welfare score.

2.2. Valuing farm animal welfare benefits with a score

The valuation of farm animal welfare in terms of changes to husbandry practices has been common practice in the literature. Examples of husbandry practices under valuation include a ban of battery cages in egg production (Bennett, 1997); changes in stocking density, ventilation, period of darkness, and percentage of flocks failing food pad lesion standards in broiler production (McVittie et al., 2005); changes to the housing system, castration techniques, tail docking and restraint in pig production (Lagerkvist et al., 2006); or changes in fodder, outdoor production, transport, and the choice of breed in livestock production (Carlsson et al., 2005).

Valuing farm animal welfare in this way is fraught with difficulties. Firstly, it does not take into account the fact that valuations of husbandry practices change as people's preferences change, or as technology finds new welfare improved husbandry techniques (McInerney, 1994). For example, as alternative husbandry practices become available, people may find existing practices less acceptable and, as a result, their valuations may change (Bennett, 1995). Thus, it would be useful if the valuation does not have to be done afresh every time science finds improved husbandry techniques or makes existing ones less expensive. Secondly, focus group discussions in several European countries have revealed a general lack of knowledge of contemporary farming practices among citizens (Miele, 2010). For example, confined systems of production were perceived to be inherently detrimental to the welfare of farm animals. Similarly, there was lack of understanding of the welfare problems of animals living in intensive production systems. All indoor systems of production were perceived to be equally detrimental to animal welfare, while all outdoor systems of production were perceived to be inherently animal friendly (Miele, 2010).

Given that farm animal welfare is a multidimensional concept that comprises physical and mental health, and aspects such as physical comfort, absence of hunger, diseases, or injuries (Veissier and Evans, 2007), and given that the amount of explanation that can be given to respondents in a valuation study is fairly limited, it would be useful if change in farm animal welfare could be presented in a succinct way. Thirdly, choice experiments on farm animal welfare have asked respondents to trade-off, for example, different levels of stocking density measured in $m^2/100kg$ (Liljenstolpe, 2008); or periods of ventilation and darkness (McVittie et al., 2005). These are arguably cognitively demanding tasks and a layperson may not have the knowledge to attempt them. They may add to the cognitive problems to which choice experiments are prone due to their repetitive nature and respondents having to evaluate large and complex choice sets (DeShazo and Fermo, 2002).

The welfare score used in this study attempts to alleviate the above mentioned problems. It allows the succinct presentation of different welfare states or welfare outcomes. It does not require respondents to be taught about animal husbandry methods that can be used to achieve particular scores. Consultations with the general public have been an important aspect of the development of the Welfare Quality® Index (Welfare Quality, 2009). This ensures that the welfare index not only reflects animal scientists' views on animal welfare but also those of the general public as it is the latter who eventually are going to use the welfare index should it be implemented, for example as a product label. A welfare score further reflects people's unwillingness to break farm animal welfare down into, what they consider to be, artificial component parts (Veissier and Evans, 2007). Finally, once the value estimates of different welfare levels are known, policy makers in collaboration with the livestock industry can decide how to achieve these welfare scores given existing technologies and animal breeds.

The welfare score is an ordinal measure which enables assignment of a score to every possible welfare state that the animals can experience such that a higher welfare state is assigned a higher score than a lower welfare state. It thus ranks animal welfare where relative welfare scores provide a ranking of the animals' well being. The value of a welfare increment using the score is derived from an individual's mapping of a score onto a perceived welfare state. Estimating WTP for different welfare score increments, we do not assume a linear relationship between respondents' utility and welfare score increments and rather expect to find relatively large increases in utility for welfare scores close to current legal minimum standards and small changes in utility when the welfare score is high. The mapping of the welfare scores on perceived welfare states may significantly differ between individuals because the information they use is incomplete and because

they understand the descriptions of the welfare scores differently. Thus, in addition to having different valuations of farm animal welfare respondents also have different perceptions of it. The latter arguably also applies to the valuation of husbandry practices as has been done in the literature with respondents mapping the effects of husbandry practices, as they understand and imagine them, onto a perceived animal welfare state that is the result of the introduction or change in a husbandry practice.

3. Methods

3.1. The welfare score

The welfare score used in this study is based on the welfare index developed by the Welfare Quality® Project. A distinctive feature of the Welfare Quality® Index is that it predominantly uses animal-based welfare outcome parameters, rather than environment-based ones. The former have the advantage that they allow comparison of the welfare of animals on different farms with different husbandry systems which makes it a versatile welfare assessment scheme. The Welfare Quality® Index is composed of about 30 measures (e.g. body condition, plumage cleanliness, behaviour). The values obtained for the different measures on a farm are combined to calculate criterion scores that reflect the compliance of the farm to twelve different criteria such as absence of injury, thermal comfort or ease of movement. This compliance is expressed on a 0 to 100 value scale where 0 corresponds to the worst situation one can find on an animal unit, 50 corresponds to a neutral situation and 100 corresponds to the best situation one can find on a farm that is the situation in which it is considered there cannot be further improvements in welfare. Because the total number of measures, the scale on which they are expressed and the relative importance of measures vary between and within criteria and also between animal types, there are three main types of calculation: decision trees, weighted sums or comparisons to alarm thresholds (INRA, 2011). The twelve welfare criteria are then aggregated into four principles according to how they are experienced by the animals (Veissier and Evans, 2007), namely good feeding, good housing, good health, and appropriate behaviour. For example, the scores obtained by a farm for criteria such as absence of injuries or absence of disease are combined to reflect compliance of this farm with the principle 'good health'. Low criterion scores are assigned more importance than high criterion scores. This ensures that farmers cannot simply compensate a low criterion score in one subcriterion with a high criterion score in another subcriterion and thus they are encouraged to correct the more severe welfare problems first. Moreover, aggregation of criteria to create an overall assessment is performed using comparisons with pre-set profiles which further limits compensation between the criteria. Hence, the higher the

aggregation is in the hierarchical structure the more limited the compensation between components (Botreau et al., 2007b). Finally, the scores obtained by a farm for all of the welfare principles are used to assign that farm to one of four welfare categories. The categories are 'Excellent' (the welfare of the animals is of the highest level), 'Enhanced' (the welfare of animals is good), 'Acceptable' (the welfare of animals is above or meets minimal requirements) and 'Not classified' (the welfare of animals is low and considered unacceptable). A farm is considered excellent if it scores more than 55 on all principles and more than 80 on two of them while it is considered enhanced if it scores more than 20 on all principles and more than 55 on two of them. Farms with acceptable levels of animal welfare score more than 10 on all principles and more than 20 on three of them. Farms that do not reach these minimum standards are not classified (INRA, 2011). For a detailed exposition of how animal welfare is measured and how scores are calculated in the Welfare Quality® Index the reader is referred to Veissier et al. (2007) and INRA (2011).

To ensure that the presentation of the welfare score alongside the contingent valuation questionnaire were comprehensible and meaningful to respondents, they were tested in three rounds of focus group discussions. The welfare score was further tested in a pilot study. In its final version it was presented on a vertical scale with short descriptions annotated at welfare levels 100, 70 and 40. The score ranged from 0 (extreme suffering) to 100 (highest possible attainable welfare). This is different to the Welfare Quality® Index which classifies farms into one of four welfare classes. The Welfare Quality® welfare classes constitute substantial changes in farm animal welfare level which can only be achieved through major changes to existing husbandry practices. In reality, any change in husbandry practices tends to improve animal welfare conditions in one specific area, such as stocking density or cage size, which often constitutes only marginal improvement in terms of overall animal welfare level over the entire life of the animals. For this reason the score in this study is on a scale of 100, comparable to those of the four welfare principles. Our score can be related to the Welfare Quality® welfare classes as depicted in Table 1. Welfare score 40 was defined as the legal minimum. In addition to the description of the

Table 1: Translation of Welfare Quality® Index classes into the welfare score

Welfare Quality® class	Welfare score
3 "Excellent"	80-100
2 "Enhanced"	60-79
1 "Acceptable"	40-59
0 "Not classified"	0-39

welfare scores, respondents were told the numbers of pigs, beef cattle and broilers that are slaughtered in the UK each year so as to give an indication of the magnitude of animals affected by welfare changes.

3.2. The contingent valuation study

The first part of the questionnaire contained a set of attitudinal questions regarding the perceived effects of high animal welfare standards on taste, safety, healthiness, and nutritional value of meat; and respondents were further asked about their households' meat consumption habits. After completing the valuation task, respondents were asked to provide information about their socio-demographic characteristics. There were two questionnaire versions containing two valuation questions each. The first questionnaire version elicited WTP for meat whose welfare score had been improved from the legislative baseline of 40 to welfare scores 60 and 80, and the second version did so for welfare scores 70 and 90. The bid amounts used and presented in Table 2 were based on the results of an open-ended contingent valuation study carried out locally, and on the results of three focus group discussions. The valuation questions were posed in the multiple-bounded uncertainty choice format (Welsh and Poe, 1998). This question format is a combination

Table 2: Bid levels

Questionnaire	Welfare score	Increase in meat expenditure per month (per year) in GBP
1	60	5 (60), 11 (132), 22 (264)
	80	8 (96), 16 (192), 32 (384)
2	70	6 (72), 12 (144), 24 (288)
	90	9 (108), 18 (216), 36 (432)

of a payment card and the polychotomous choice question introduced by Ready and Hu (1995). Respondents were given a panel of $k = 3$ bid amounts. At each bid amount, they were asked to give their subjective assessment of the probability of them accepting the bid in the form of verbal statements ranging from 'Definitely yes' (Y), 'Probably yes' (PY), 'Don't know' (DK), 'Probably no' (PN) to 'Definitely no' (N). Verbal statements were used instead of numerical expressions because respondents tend to be poor at responding to or stating probabilities in numerical terms (Evans et al., 2003). We expected respondents to be uncertain about their preferences because in Britain the only animal welfare label is RSPCA Freedom Food which accounts for only around 2% of the livestock product market in the UK (Mayfield et al., 2007) and therefore respondents were expected to have little experience in purchasing meat with animal welfare credentials. Other reasons

9

why respondents may be uncertain about their prefererences that have been given in the literature include respondents having incomplete knowledge about their true valuation (Li and Mattsson, 1995); or respondents having an implicit valuation distribution rather than a single true point value in mind (Wang, 1997).

Data collection was carried out in June 2009 using the phone-mail/email-phone method which involved two rounds of phone calls and a mail/email component. Respondents were contacted in a first round of phone calls. After agreeing to participate in the survey they received the questionnaire alongside the information sheet about the welfare score through the mail, or they were sent a link to access this information via email. Subsequently, they were telephoned to supply their answers. In total, 2719 people were contacted using Random Digit Dialling and 490 people recruited to achieve 300 interviews. In the initial phone conversation respondents were contacted in a first round of phone calls. To avoid self-selection of people particularly concerned about animal welfare, respondents were told that the topic of the valuation study was to do with farming and meat. The sample was stratified according to age and socio-economic group. In total, 278 respondents with meat consuming households were included in the analysis.

3.3. Econometric specification

The ordered probit model has been used in various studies to analyse multiple bounded uncertainty choice format data (Wang, 1997; Alberini et al., 2003; Balcombe et al., 2009b). It is assumed that the unobserved continuous dependent variable is the respondent's true valuation of a change in animal welfare scores. Stated preference studies using this approach of parameterising the model in WTP space have been carried out by Cameron and James (1987), Train and Weeks (2005), Scarpa et al. (2008), and Balcombe et al. (2009a). Thus, the latent utility of respondent i that cannot be observed by the researcher is specified as

$$u_i = \mu + \alpha' z_i - b + e_i \sigma \tag{1}$$

where μ is the constant representing mean utility; z_i are individual characteristics which are given zero mean by construction; b is the bid; e is the error term distributed standard normal; and σ is the standard deviation of the error. Ordering among the latent variables was introduced through the observed choices of respondents in the following way

'No'	$y_i = 1$	if $u_i < 0$
'Probably No'	$y_i = 2$	if $0 < u_i < \lambda_1$
'Don't know'	$y_i = 3$	if $\lambda_1 < u_i < \lambda_2$
'Probably Yes'	$y_i = 4$	if $\lambda_2 < u_i < \lambda_3$
'Yes'	$y_i = 5$	if $\lambda_3 < u_i < \lambda_4$

The thresholds, λ, at which respondents switch among the five qualitative response categories are also unobserved by the researcher. They were estimated as distances from the mean of the respondent's value distribution. If a respondent chose 'Don't know' her probability lay between the following thresholds

$$\frac{\lambda_1 - \lambda_2}{2} < u_i^\star < +\frac{\lambda_2 - \lambda_1}{2} \tag{2}$$

where $u_i^\star$ arises because (1) is consistent with the assumption that: $u_i = u_i^\star + \frac{\lambda_1 + \lambda_2}{2}$. The underlying utility model then is :

$$u_i^\star = \gamma + \alpha' z_i - b + e_i \tag{3}$$

where z_i has zero mean and

$$\gamma = \mu - \frac{\lambda_1 + \lambda_2}{2} \tag{4}$$

is the mean WTP. Identification is achieved by setting $\lambda_0 = -\infty$, $\lambda_1 = 0$ and $\lambda_J = \infty$, and by restricting the coefficient for the bid to -1. While it is common practice to fix the variance in the ordered probit model to one and calculate WTP as a ratio of two random variables causing the WTP distribution to have an infinite variance, the present study restricts the bid coefficient so that the intercept, μ, in (1) can be interpreted as the WTP distribution with a finite variance. Using Bayesian inference the likelihood is combined with a normal prior for β and a gamma distributed prior for the precision, h. Posterior estimation was carried out using Gibbs sampling with data augmentation (Albert and Chib, 1993). After a burn-in of 20,000 draws, there were 200,000 iterations with every 20th value being sampled, leaving 10,000 draws to be analysed. Convergence of the sampler was monitored visually and by conducting a modified t-test for the hypothesis of 'no-difference' between the first and second half of the sampled values on the sequences of parameters. WTP estimates for each welfare increment was obtained from the baseline models which did not include any explanatory variables.

A second model was estimated including a set of socio-demographic variables, and a

third model including a set of attitudinal variables. It was expected that WTP for animal welfare scores increases with education and income levels and decreases with age. Moreover, given that high welfare meat is perceived to be healthier, it was expected that women who tend to be more health concious would have a higher willingness to pay. In the third model, the attitudinal variables were included as dummy variables with a value of one if a respondent strongly agreed or agreed with the notions that high welfare meat is more nutritious, safer, and better for the environment, respectively, and zero otherwise. It was expected that agreement with these statements would have a positive effect on WTP.

4. Results and discussion

The data was collected in 2009. Total sample size was $n = 300$. The sample stratification was as follows. Minimum quotas to account for 200 interviews were set in terms of gender (100 male and 100 female); socio-economic-group (AB:50, C1:60, C2:40, DE:50) 1; and age (18-39: 70, 40-59: 70, 60+: 60). These quotas were set to ensure that the sample population would be representative of the UK population. The remaining 100 interviews were let to fall out naturally, with the restriction of a maximum of 90 interviews for the over sixties for the entire sample. This restriction ensured that elderly people would not be overrepresented as they were expected to be most likely to be willing to take part. Since 28% of the population are expected to be in the over 60 age group, which would be around 85 people out of 300, the restriction is not expected to bias the sample. Average age in the sample was 50.45 years, and mean income before tax was £48,9766. Hence, the average income of respondents who stated their income was above average. To compare, according to HM Revenue and Customs' data from 2004-2005, mean income before tax was £22,800 per year. The vast majority of respondents was white (95.0%) which is only a slight overrepresentation of whites as according to the 2001 census 92.1% of people in the UK were white. Sixty one percent of respondents were female, most likely due to the fact that the primary household shopper was requested to complete the survey. Average weekly meat expenditure was £17.02. Given inflation, this broadly compares with the British average of £10.10 in 2005/06 reported by the Office of National Statistics.

Answers to the attitudinal questions revealed that respondents 'strongly agreed' or 'agreed' with the notion that meat from animals with high welfare is healthier (78%); better tasting (71.7%); has better nutritional value (72%); has better food safety (75%); and is better for the environment (72.3%). The fact that consumers perceived high welfare meat as distinct from conventional products suggests that these products may be able to command a price premium in the market. Furthermore, the majority of respondents (94.3%) stated

that they would like to see a welfare scoring system to be used in food stores indicating support for the introduction of a labelling scheme. It appears that an important aspect in the valuation of meat with high welfare scores is its perceived better quality. Provided that the latter could be achieved through other means, respondents' premium for meat with high welfare scores may be lower. Still, 75.7% of respondents 'strongly agreed' or 'agreed' with the notion of being concerned about the way farm animals are treated which may suggest that animal welfare itself is an important component in their valuation of meat with high welfare scores. Moreover, agreeing with the notion that meat from high welfare animals has better quality, has only a limited effect on WTP as discussed below and depicted in Table 5.

Table 3: Results of the basic ordered probit model

Version	Welfare Score	μ	
		Mean	Std. Dev.
1	60	28.581	1.636
	80	37.183	1.928
2	70	26.882	1.505
	90	31.747	1.597

Table 4: Monthly WTP in terms of an increase in absolute meat expenditure (in £) and as a proportion of average monthly meat expenditure

Version	Welfare score	WTP	in %	n
1	60	19.36	26	142
	80	23.65	32	
2	70	18.77	30	136
	90	21.20	34	

Results from the ordered probit model found that all coefficients had the expected signs and their 95% Bayesian credible intervals excluded zero [1]. The mean WTP estimates reported in Table 4 show that respondents in the first subsample were willing to pay an additional £19.36 and £23.65 per month to ensure that their meat was produced

---

[1]The 95% Bayesian confidence interval is given by the posterior mean $\pm$ 1.96 $\times$ posterior standard deviation

to welfare scores 60 and 80, respectively. Respondents in the second subsample were willing to pay an additional £18.77 and £21.20 per month for their meat to be produced to welfare scores 70 and 90, respectively.

When comparing WTP estimates for the different welfare scores across subsamples it has to be noted that respondents were asked to state their WTP in terms of an increase in their absolute monthly meat expenditure. The mean WTPs therefore have to be considered in relation to the average monthly meat expenditure of each subsample which was £73.28 in the first subsample and £62.66 in the second subsample. The lower average meat expenditure in the second subsample is most likely due to its lower average household income indicated by one of twelve income categories. While the first subsample had an average income category membership of 6.9, it was slightly lower in the second subsample, namely 6.7. WTP estimates between subsamples can still be compared in terms of the percentage of respondents' average meat expenditure. These are reported in the fourth column of Table 4. They show that respondents were willing to increase their monthly meat expenditure by 26% and 32% for meat produced to welfare scores 60 and 80, respectively, and to increase their monthly meat expenditure by 30% and 34% for meat produced to welfare scores 70 and 90, respectively. The fact that respondents are willing to increase their meat expenditure by about a third gives some indication of the magnitude of the price premium that labelling of higher welfare meat might be able to achieve. It can be expected that higher welfare scores are more desirable to respondents. In our study this translates into higher WTP for larger improvements in welfare scores. The legal minimum welfare score of 40 and thus the perception of the current farm animal welfare situation in the UK is likely to affect utility of additional improvements in welfare scores such that WTP for incremental changes in animal welfare scores is relatively lower if it exceeds satiation where additional units of welfare scores above the satiation level have zero marginal utility. We find evidence of satiation in the case of WTP for improving welfare scores from 40 to 80 and to 90, respectively. This means that consumers may form their preferences for a particular welfare score based on its relative position on the scale. The implication of this for labelling of high welfare meat is that higher welfare scores that exceed satiation levels for animal welfare, which are more expensive to implement and therefore would require a larger price premium, may not be able to command correspondingly higher price premia.

Another issue that warrants further investigation in our study is the fact that WTP as a percentage of meat expenditure changes by about 2% between welfare scores which may seem relatively small. However, if the price premium is passed on directly to the

farmer this would constitute a substantial increase in producer price. Thus, it appears that currently British consumers do have sufficiently distinct preferences for different animal welfare levels that would allow rewarding farmers for their extra effort. In addition, given that we used a very simplistic welfare score and still managed to communicate differences in welfare standards which resulted in relatively different WTP amounts shows that a label in the form of a score may be able to communicate the differences in animal welfare outcomes and achieve accordingly differentiated price premia.

The results of the estimation of the ordered probit model with a number of socio-demographic variables and with a number of attitudinal variables are reported in Table 5. Apart from being a woman, none of the socio-demographic variables was found to have

Table 5: Determinants of WTP				
Welfare score	60	70	80	90
income	5.518	10.452	13.434	11.281
	(71.368)	(71.105)	(71.182)	(72.250)
female	245.032*	175.922*	223.63*	142.754*
	(56.515)	(52.494)	(52.735)	(70.627)
age	-19.464	8.746	-28.841	-10.033
	(17.181)	(13.689)	(18.607)	(23.360)
education	13.368	19.751	12.704	27.447
	(17.388)	(16.477)	(18.043)	(27.447)
high welfare meat is ...				
...more nutritious	-0.201	1.831	1.225	3.261
	(70.158)	(70.323)	(71.095)	(71.273)
...safer	7.244*	2.858	9.670*	1.212
	(2.358)	(1.914)	(3.553)	(2.601)
..better for the environment	6.950*	0.014	10.934*	3.623
	(1.974)	(2.190)	(2.947)	(2.981)
We have moral obligation to	20.939*	23.989*	25.863*	26.109*
safeguard animal welfare	(2.342)	(2.356)	(3.300)	(2.845)

* The 95% Bayesian confidence interval excluded zero.

a strong impact on WTP. By contrast, agreeing with the notion that we have a moral responsibility to safeguard the welfare of animals had a substantial and positive effect on WTP for high welfare status meat. Agreement with the notion that high welfare meat is safer had no significant effect on WTP, and agreement with the notion that high

welfare meat is safer had a positive effect on WTP only in the case of one subsample. Generally, it appears that attitudes may be better at explaining preference heterogeneity than socio-demographics. Following the notion of O'Connor et al. (1999), this may be due to the fact that farm animal welfare has become such a consensual issue that social characteristics are no longer of great importance and differences are rather found at the level of attitude-based distinction.

A limitation of this research is with regard to the reliability and validity of its results. Since no real money was involved in the surveys, the research was not incentive compatible, which may have resulted in respondents overstating their WTP. Respondents' choices and answers to the WTP questions may not have been expressions of preferences which are comparisons between alternative states of the world but expressions of attitudes (Kahneman and Sugden, 2005). In addition, animal welfare is a very emotive subject which may have lead respondents to further overstate their WTP. This hypothetical bias has to be acknowledged, although the size of any such bias has not been assessed. Finally, this study valued animal welfare scores using an ordinal scale. Admittedly, there are various problems associated with the valuation of an ordinal score. The value of a welfare increment using the welfare score is derived from an individual's mapping of a score onto a perceived animal welfare state. Providing a more detailed description of three given welfare scores (40,70,90) and therefore anchoring scores onto welfare states by descriptions may assist people in performing this mapping but individuals' mappings are still likely to differ across the population because the information they use is incomplete and because they understand the descriptions differently. Accordingly, when valuing incremental changes in the welfare score, people may not only be valuing welfare differently, but valuing different perceptions of welfare.

Still, choices involving uncertainty and ambiguity are common in the real market place. Product labels such as the traffic light colour coding for food in Britain are already in place and product information in the form of a welfare score is likely to be introduced in the EU in the near future. Respondents are therefore likely to become more accustomed to the concept of farm animal welfare scores and their reactions in the focus groups and respondents' comments to debriefing questions suggest that a welfare score would be well accepted by the general public as a representation of animal welfare. Moreover, once the Welfare Quality® Index becomes operational in the EU, respondents are likely to become more accustomed to considering different animal welfare scores.

5. Conclusion

Policy makers are considering the introduction of an EU animal welfare label which may take the form of a certified grade or a score indicating the welfare level to which products have been produced. For a welfare labelling scheme to be viable and desirable, the market for livestock products has to be sufficiently segmented with consumers having sufficiently distinct preferences for products produced to different welfare standards. The present study investigates the preferences of British consumers for meat produced to different welfare levels. In doing so a hypothetical welfare score is used which is based on the Welfare Quality® Index, a recently developed on-farm welfare assessment scheme. This is a new approach of presenting welfare outcomes to respondents when valuing farm animal welfare benefits which have been valued to date in terms of changes to livestock husbandry practices. We find that meat produced to higher welfare standards is perceived by respondents as having better quality and WTP for it tends to be higher. Respondents are willing to pay an increasingly larger proportion of their meat expenditure for meat with higher welfare scores. We conclude that given the level of differentiation of people's preferences for different levels of farm animal welfare, the introduction of a labelling scheme in the form of a certified logo or a rating system appears generally feasible. WTP is affected by respondents' interpretation of the welfare score and their perception of the current farm animal welfare situation in the UK. The latter appears to affect WTP for improvements in welfare scores with WTP at higher welfare scores being relatively low suggesting that welfare labels that appear to indicate welfare standards that exceed consumers' satiation levels for farm animal welfare may not be able to command a sufficiently high price premium, depending on what the actual costs of achieving higher welfare scores are. The main food policy implications of this study are that consumers who have preferences for high welfare meat currently cannot satisfy them because of a lack of information. Using an outcome-based farm animal welfare assessment scheme and labelling products with certified logo or a welfare score accordingly could address this to the benefit of consumers, the farm industry and society in general.

References

Akerlof, G.A., 1970. The market for 'lemons': Quality uncertainty and the market mechanism. Quarterly Journal of Economics 84, 488–500.

Alberini, A., Boyle, K., Welsh, M.P., 2003. Analysis of contingent valuation data with multiple bids and response options allowing respondents to express uncertainty. Journal of Environmental Economics and Management 45, 40–62.

Albert, J.H., Chib, S., 1993. Bayesian analysis of binary and polychotomous response data. Journal of the American Statistical Association 88, 669–679.

Balcombe, K.G., Chalak, A., Fraser, I., 2009a. Model selection for the Bayesian mixed logit. Journal of Environmental Economics and Management 57, 226–237.

Balcombe, K.G., Fraser, I., Samuel, A., 2009b. Estimating WTP with uncertainty choice contingent valuation. Paper provided by Department of Economics, University of Kent - KDPE 09/21.

Bennett, R.M., 1995. The value of farm animal welfare. Journal of Agricultural Economics 46, 46–60.

Bennett, R.M., 1997. Farm animal welfare and food policy. Food policy 22, 281–288.

Botreau, R., Bonde, M., Butterworth, A., Perny, P., Bracke, M.B.M., Capdeville, J., Veissier, I., 2007a. Aggregation of measures to produce an overall assessment of animal welfare. Part 1: A review of existing methods. Animal 1:8, 1179–1187.

Botreau, R., Bracke, M.B.M., Perny, P., Butterworth, A., Capdeville, J., van Reenen, C.G., Veissier, I., 2007b. Aggregation of measures to produce an overall assessment of animal welfare. part 2: analysis of constraints. Animal 1:8, 1188–1197.

Cameron, T.A., James, M.D., 1987. Estimating willingess to pay from survey data: an alternative pre-test-market evaluation procedure. Journal of Marketing Research 24, 389–395.

Carlsson, F., Frykblom, P., Lagerkvist, C.J., 2005. Consumer preferences for food product quality attributes from Swedish agriculture. Ambio 34, 366–370.

Caswell, J.A., 1998. How labeling of safety and process attributes affects markets for food. Agricultural and Resource Economics Review 27, 151–158.

Darby, M.R., Karni, E., 1973. Free competition and the optimal amount of fraud. Journal of Law and Economics 16, 67–88.

Deimel, I., Franz, A., Frentrup, M., von Meyer, M., Spiller, A., Theuvsen, L., 2010. Perspektiven für ein Europäisches Tierschutzlabel. BLE Projektkennzifer 08HS010.

DeShazo, J.R., Fermo, G., 2002. Designing choice sets for stated preference methods: The effects of complexity on choice consistency. Journal of Environmental Economics and Management 44, 123–143.

European Commission, 2007. Attitudes of EU citizens towards animal welfare. Special EUROBAROMETER 270 .

Evans, M.F., Flores, N.E., Boyle, K.J., 2003. Multiple-bounded uncertainty choice data as probabilistic intentions. Land Economics 79, 549–560.

FAWC, 2006. Report on welfare labelling. Farm Animal Welfare Council.

INRA, 2011. Calculation of scores. www1.clermont.inra.fr/wq/protocol_calculation_scores.php.

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

Lagerkvist, C.J., Carlsson, F., Viske, D., 2006. Swedish consumer preferences for animal welfare and biotech: A choice experiment. AgBio Forum 9, 51–58.

Li, C.Z., Mattsson, L., 1995. Discrete choice under preference uncertainty: an improved structural model for contingent valuation. Journal of Environmental Economics and Management 28, 256–269.

Liljenstolpe, C., 2008. Evaluating animal welfare with choice experiments: an application to Swedish pig production. Agribusiness 24, 67–84.

Lusk, J.L., Nilsson, T., Foster, K., 2007. Public preferences and private choices: effect of altruism and free riding on demand for environmentally certified pork. Environmental and Resource Economics 36, 499–521.

Mayfield, L.E., Bennett, R.M., Tranter, R.B., Wooldridge, M.J., 2007. Consumption of welfare-friendly food products in Great Britain, Italy and Sweden, and how it may be influenced by consumer attitudes to, and behaviour towards, animal welfare attributes. International Journal of Sociology of Food and Agriculture 15, 59–73.

McInerney, J., 1994. Animal welfare: An economic perspective, in: Bennett, R.M. (Ed.), Valuing farm animal welfare, University of Reading, Workshop held at the University of Reading, 30 September 1993. pp. 9–25.

McVittie, A., Moran, D., Sanilands, V., Sparks, N., 2005. Estimating non-market benefits of reduced stocking density and other welfare increasing measures for meat chickens in England. Final report SAC Commercial Ltd. Project no. AW0236.

Miele, M., 2010. Report concerning consumer perceptions and attitudes towards farm animal welfare. Uppsala University, Official deliverable of the EU VII Framework project European Animal Welfare Platform.

Nocella, G., Hubbard, L., Scarpa, R., 2010. Farm animal welfare, consumer willingness to pay, and trust: Results of a cross-national survey. Applied Economic Perspectives and Policy 32, 275–297.

O'Connor, R.E., Bord, R.J., Fisher, A., 1999. Risk perceptions, general environmental beliefs, and willingness to address climate change. Risk analysis 19, 461–471.

Ready, R.C., Hu, D., 1995. Statistical approaches to the fat tail problem for dichotomous choice contingent valuation. Land Economics 71, 491–499.

RSPCA, 2010. Record-breaking demand for freedom food. http://www.rspca.org.uk/media/news (visited 15/01/2011).

Scarpa, R., Thiene, M., Train, K., 2008. Utility in willingness to pay space: a tool to address confounding random scale effect in destination choice to the alps. American Journal of Agricultural Economics 90, 994–1010.

Train, K., Weeks, M., 2005. Discrete choice models in preference space and willingness-to-pay space, in: Scarpa, R., Alberini, A. (Eds.), Applications of simulation methods in environmental and resource economics. Springer, Dordrecht, The Netherlands. volume 6 of *The economics of non-market goods and resources*, pp. 1–16.

Veissier, I., Evans, A., 2007. Rationale behind the Welfare Quality®assessment of animal welfare, in: Veissier, I., Forkman, B., Jones, B. (Eds.), Second Welfare Quality®stakeholder conference "Assuring animal welfare: from societal concerns to implementation", 3-4 May 2007, Berlin, Germany.

Veissier, I., Forkman, B., Jones, B. (Eds.), 2007. Assuring animal welfare: from societal concerns to implementation - Second Welfare Quality®stakeholder conference, Berlin.

Wang, H., 1997. Treatment of "don't-know" responses in contingent valuation surveys: A random valuation model. Journal of Environmental Economics and Management 32, 219–232.

Welfare Quality, R., 2009. http://www.welfarequality.net/everyone.

Welsh, M.P., Poe, G., 1998. Elicitation effects in contingent valuation: Comparisons to a multiple bounded discrete choice approach. Journal of Environmental Economics and Management 36, 170–185.