The benefits of farm animal welfare legislation: the case of the EU broiler directive and truthful reporting

Article

Accepted Version

The benefits of farm animal welfare legislation: the case of the EU broiler directive and truthful reporting

Richard Bennett, Kelvin Balcombe, Philip Jones and Andrew Butterworth[1]

[Original submitted October 2016, Revisions received February 2017 to January 2018, Accepted February 2018]

ABSTRACT

The EU broiler Directive came into force in the UK in June 2010 with the aim of setting new minimum standards, monitoring broiler welfare and addressing any welfare problems. A survey questionnaire was used to elicit information from a stratified sample of citizens in England and Wales regarding their willingness to pay for the provisions of the Directive, as an estimate of the consumer surplus associated with the legislation. We also explore the usefulness of Prelec's (2004) Bayesian Truth Serum (BTS) in promoting respondents' truthful reporting. A median willingness to pay of £21.5 per household per year (corrected for sample bias and possible 'yea saying') was estimated from 665 responses. This provides an estimated benefit of the legislation to citizens of over £503 million/yr, equivalent to 5.3% of current consumer expenditure on chicken. This compares to an estimated £22 million per annum cost of producers' compliance and government enforcement associated with the legislation. No statistically significant differences in responses between respondents that did and did not have a BTS incentive to answer questions truthfully were found, which might reflect apparently truthful answers in this case, an insufficiently strong financial incentive or a weakened effect due to an element of disbelief in the BTS amongst the sample. The analysis suggests that the benefits of the broiler Directive to citizens greatly outweigh the additional costs to producers, making a case for the legislation to be retained.

[1] Richard Bennett (contact author: E-mail r.m.bennett@reading.ac.uk) Kelvin Balcombe and Philip Jones are all in the School of Agriculture, Policy and Development, University of Reading, UK. Andrew Butterworth is in the Department of Clinical Veterinary Science, University of Bristol, UK. The authors would like to thank the Department for Environment, Food and Rural Affairs for funding the study to evaluate the effectiveness of implementation of the EU Broiler Directive in England and Wales, of which the research reported here formed a part. We would also like to thank the producers and citizens who participated in the study. In addition, our thanks are due to two anonymous reviewers and the Editor for comments on earlier drafts of this paper. Any errors or omissions are the responsibility of the authors.

KEYWORDS: Animal welfare; EU broiler legislation; willingness to pay; truthful reporting; Bayesian Truth Serum

JEL Classifications: *C11, C18, D12, D46*

The benefits of farm animal welfare legislation: the case of the EU broiler directive and truthful reporting

1. Introduction

Legislation has been a major policy instrument for the protection of animal welfare in Europe with many European Union (EU) regulations and directives implemented in the last 20 years (Bennett and Appleby, 2011; European Commission, 2016). In 2007 new EU rules were agreed (Council Directive 2007/43/EC) for protecting the welfare of broiler chickens (European Commission, 2007). The broiler Directive came into force in the United Kingdom (UK) on 30 June 2010. When such legislation is introduced there is an expectation that the benefits of that legislation will be greater than the costs. Expected costs include costs to government, for administration, monitoring and enforcement, as well as compliance costs for producers and broader industry and national costs associated with impacts on competitiveness, international trade and economic output. Expected benefits of such legislation are generally non-monetary, but include consumer perceptions of enhanced chicken meat quality and enhancement of consumer and citizens levels of satisfaction from knowing that the welfare of broiler chickens is better protected in the food production process (see Mayfield *et al.*, 2007 in relation to consumer attitudes to animal welfare in Europe). Indeed, a UK government *ex ante* Impact Assessment (IA) published in 2010 (Defra, 2010) concluded that while the broiler Directive would lead to some costs to producers, this would be more than off-set by non-market benefits to consumers, in the form of higher broiler welfare, for which many would be willing to pay. We test the hypothesis that citizens have an additional willingness to pay (WTP) for the broiler Directive beyond any price increase as a result of increased costs of production due to the legislation and assess the scale of such benefits by means of an *ex post* WTP survey of citizens. Since the perceived costs of improving animal welfare can deter legislation it is important to estimate the benefits of legislation to better inform policy. In addition, as a methodological contribution, we explore the usefulness of Prelec's (2004) Bayesian Truth Serum (BTS) which is designed to reduce respondents' misreporting (i.e. help ensure 'truthful' reporting) in surveys.

Theoretically, the WTP we estimate using the CV method is a Hicksian consumer surplus measure (see Mitchell and Carson, 1989 p25) which can be interpreted as a respondent's maximum WTP for the broiler legislation and its provisions (considered, at least in part, as a public good, because some of the benefits accruing from the legislation are non-excludable and non-divisible/non-rival).

Our WTP estimate measures both use and non-use value and, as noted by Zhao and Kling (2004), is defensible theoretically as a welfare measure that can be directly applied to cost-benefit analysis.

Section 2 covers the background of WTP in the area of animal health and welfare and associated legislation along with a description of the broiler Directive and its implementation in the United Kingdom. Section 3 outlines the WTP survey whilst Section 4 covers the analytical methods. Section 5 presents the results and Section 6 discusses them. Section 7 offers some conclusions.

2. Background

WTP applied to animal health and welfare

Although widely used in other areas, stated preference valuation methods have perhaps been less widely applied to animal welfare. Bennett *et al.* (2011) provide a review and critique of valuation studies applied to animal welfare. Lagerkvist and Hess (2011) identify 24 stated preference studies of people's WTP in relation to animal welfare, yielding 106 WTP estimates, in their meta-analysis. Approximately half of these are contingent valuation studies and half use a choice experiment approach (with one other using an experimental auction method). Additionally, Clark *et al.* (2016) observed just 17 WTP studies out of some 80 studies included in their systematic review of public attitudes, perceptions and behaviours towards animal welfare concerns arising from livestock production diseases.

A common feature of these studies is that they elicit WTP from citizens regarding specific changes in husbandry or other practices to improve welfare, although Bennett *et al.* (2012) present a method for the economic valuation of animal welfare benefits more generally using a single welfare score. WTP values can be used as estimates of the likely magnitude of the benefits that citizens obtain from each of these animal husbandry or other measures. It is clear from a number of the studies reviewed that consumers value animal welfare not only for ethical reasons but also in some cases because they believe that products from animals with higher welfare are of higher quality in terms of taste, nutrition and safety and better for the environment (despite there being little empirical evidence to support the validity of these beliefs).

The broiler Directive and its implementation in the UK

While one of the purposes of the broiler Directive is to promote more universal achievement of minimum animal welfare standards across EU members states (MS), a number of implementation options are permitted, including variation in the maximum stocking density limits that can be applied in each MS. In the UK producers have been permitted to choose from among two stocking density maxima:

i) stocking up to 33kg live weight per m^2 is permitted if specific standards are met for drinkers, feeding, litter, ventilation and heating, noise, light, inspections, cleaning, record keeping, training and surgical interventions.

ii) stocking beyond 33kg per m^2 up to 38kg/m^2 is permitted if an additional set of standards are met (the Annex II requirements). These include notification and documentation requirements, plus further controls on environmental parameters in broiler housing.

In addition, the Directive requires the collection of data from farms on cumulative daily mortality (CDM) and data from slaughterhouses for eight post-mortem measures of body condition (collectively known as the 'trigger conditions') to help identify poor welfare on farms. These data, which relate to each batch of birds per farm sent to slaughter, are used by the Food Standards Agency (FSA) and the Animal and Plant Health Agency (APHA) to identify farms which may require problem notification and/or on-site inspection. The post-mortem body condition measures for which data are collected are: 1) ascites/oedema, 2) cellulitis and dermatitis, 3) dead on arrival, 4) emaciation, 5) joint lesions/arthritis, 6) septicaemia/respiratory problems, 7) total carcase rejections and 8) foot pad dermatitis score. Evaluation of these data involves two processes:

Process 1: An alert (to APHA) is triggered if the incidence of any of the individual post-mortem conditions is exceptionally high in any batch (defined as greater than six standard deviations above the mean);

Process 2: An alert is triggered if the CDM is unusually high (defined as greater than three standard deviations above the mean) and, additionally, the level of three or more of the post-mortem conditions is high (defined as above the mean).

When trigger thresholds are breached, the keeper of the animals and APHA are alerted by

means of a 'trigger report'. APHA and FSA have inspection regimes and data handling systems to communicate information relating to poor welfare between the slaughterhouse and the producer. Investigative action will be taken by APHA Veterinary Officers, and this may include requesting a written action plan to remedy the problem from producers and/or a visit to the production site. APHA may, in addition, carry out a number of random welfare inspections.

Thus, under the broiler Directive, new broiler production standards are set and enforced through a system which continuously monitors key welfare indicators, linked to an intervention process to deal with problems should they occur. In 2013/14 the percentage of batches sent for slaughter in GB which exceeded Trigger 1 levels varied from less than 0.2% for joint and arthritis problems to 1.8% for foot pad dermatitis (Food Standards Agency, personal communication).

3. The Survey

Questionnaire design and survey administration

A questionnaire was designed which contained an introduction explaining the nature of the survey and its purpose, followed by questions to respondents with regards to their:

1. Personal characteristics (sample stratification variables);
2. Current consumption of chicken meat;
3. Attitudes towards farmed-animal welfare;
4. WTP for farmed-animal welfare improvements in general;
5. Attitudes towards the broiler Directive;
6. WTP for the Directive and debriefing questions to explain respondents' WTP responses;
7. Socio-demographic characteristics (non-stratification variables).

A specimen copy of the questionnaire can be found in the online Appendix, available at the publisher's website. Two versions of the questionnaire were designed in relation to those who did, and did not, consume chicken. Chicken consumers were asked how much extra they would be willing to pay for the Directive in the form of a premium on the price of the conventionally reared chicken products that they purchased, while non-chicken eaters were asked their WTP as an

152 additional sum on their income tax. The survey was carried out by means of a web-based
153 questionnaire hosted by the commercial research support company Qualtrics
154 (http://www.qualtrics.com). The questionnaire was sent to a Qualtrics-secured panel of the general
155 public in England and Wales. A stratification procedure was applied during recruitment of survey
156 participants to ensure that the sample was broadly representative of populations in England and
157 Wales. This required set proportions across geographic area, age and income, based on population
158 distributions from data from the Office of National Statistics (ONS).[2] Respondents were screened
159 to ensure that they were the primary, or regular, purchaser of food for their household.
160
161 Respondent WTP was elicited using a contingent valuation (CV) payment card with a discrete
162 dichotomous choice format with multiple increasing values akin to a 'bidding game' (see, for
163 example, Heinzen and Bridges, 2008, who compare four different CV elicitation methods). The
164 payment card method is regarded as efficient, robust and reliable (see, for example, Bateman, *et*
165 *al*, 2002; Pearce and Ozdemiroglu, 2002), and was considered the most appropriate WTP
166 elicitation method for our purposes. We sought people's holistic WTP for a single specified
167 legislative intervention rather than valuation of different attributes of the legislation where a choice
168 experiment approach would have been more appropriate. Our hypothesis was that people would
169 have a willingness to pay in addition to current market prices of chicken meat to retain the
170 legislation up to the point where their consumer surplus is zero. After reminding respondents that
171 they have a limited household budget and that additional money spent on supporting the chicken
172 legislation may mean that they have less money to spend on other things, respondents were
173 presented with a range of seven bids, of ascending value, from '5 pence per week extra' to 'more
174 than £4 per week extra'. These bids were expressed as an additional amount that respondents would
175 pay per week, either in the form of an increase in the price of chicken meat, or taxation for those
176 relative few who did not consume chicken[3]. The range of amounts chosen was based on findings

[2] Despite this, some over-recruitment was found in certain socio-demographic groups requiring adjustment described later in the Results section. The questionnaire was pre-tested by the research team on a small number of consumers, and then formally piloted by Qualtrics using a panel of 50 consumers. The survey was officially launched on 30 September 2014 and was closed on 8 October 2014, by which time 665 usable responses had been received. In order to eliminate the possibility of partially completed questionnaires, respondents were required to complete all questions before they could file their return.

[3] A zero amount was not presented to respondents at this point because (i) respondents could answer 'No' to all the bids presented to them including the lowest 5 pence per week bid and (ii) a previous question had already

of pre-testing of the questionnaire and was later confirmed to be appropriate in the pilot. Respondents were asked to tick 'Yes' if they would be willing to pay each amount or 'No' if they would not. Respondents were asked to state their WTP to support the Directive and its associated provisions, taken as a whole. Respondents were also asked to indicate how they thought other people would respond to the WTP questions.

Misreporting and the Bayesian Truth Serum (BTS)

Prelec (2004) proposed the BTS as a general mechanism to encourage 'truth telling' in a variety of contexts, from answering simple 'yes or no' questions to more complex WTP studies. There has been growing recognition of the BTS as a potential incentive mechanism for accurate reporting across a range of survey types (e.g. John *et al*, 2012; Weaver and Prelec, 2013). Incentives to accurate reporting may be worth using where respondents believe (perhaps falsely) that there may be real consequences induced by the findings of a survey and so there is the potential for the hypothetical survey to be 'incentive incompatible' (i.e. for respondents to misreport their true preferences; see Hurwicz, 1972 p.320), and also where reporting accurately is costly in terms of time or cognitive effort or where respondents have little incentive to answer carefully and honestly.

The aim of the BTS approach is to elicit truthful (i.e. honest and carefully considered) subjective data in situations where the objective truth is unknowable. The method uses an information scoring system that is thought to elicit truthful answers from a sample of rational expected-value-maximizing respondents (Prelec, 2004). It has been claimed that the approach can eliminate bias common in contingent valuation studies when applied to responses pertaining to the respondent's contribution to a public good (Weaver and Prelec, 2013). The BTS can be described as follows. People are asked a question (e.g. would you be prepared to pay an additional amount for legislation x?). Additionally, they are asked to estimate what proportion of people (in general) they think would give a particular response. Both these responses contribute to a formula made known to the respondent, which integrates their responses with the responses of other individuals asked the same questions. The BTS theorem contends that if individuals provide personally truthful answers then expected scores are maximized. Respondents are then incentivised (usually

asked respondents whether they would be likely to be willing to pay something for the legislation on a five-point scale (Definitely yes, Possibly yes, Not sure, Probably not, Definitely not).

monetarily) to get a higher score. The precise nature of the formula is shown in the online Appendix, available at the publisher's website. The formula assigns high scores to respondents' personal answers that are more common across the sample than collectively predicted by all respondents when asked how they think others would respond (i.e. what proportion of people generally would give a particular response). For example, an answer shared by 10% of all respondents compared to a prediction of 5% would be surprisingly common and receive a high score. The assumption is that people tend to believe that the proportion of responses that correspond to their own response will be higher than the mean proportional responses elicited from the whole sample. By rewarding people for giving 'surprisingly common' responses truth telling is thus encouraged.

Half of the respondents were given a Bayesian Truth Serum Incentive (BTSI) statement to see if it had an impact on their reporting (see Weaver and Prelec, 2013 for other examples of such statements). The BTSI statement read: "Please note that one person will be chosen at random in a prize draw to receive up to £100. The exact amount of the prize will be determined by the winner's Truth Score (the higher the score the higher the prize amount). The Truth Score is a measure recently developed and published in the academic journal Science (http://nel.mit.edu/pdf/17BayesianTruthSerumcopy.pdf). Even though only you know how truthful your answers are, people who consider the questions carefully, answer honestly and take care to avoid mistakes score higher on the Truth Score and provide more reliable information for the survey." For the BTS to work, respondents are not required to understand the mathematics of the scoring nor the theory behind the BTS. However, they must believe that the method rewards truth telling on average. If the BTSI had an impact, this should manifest itself as a difference in the responses of incentivised vs non-incentivised individuals.

4. Estimation and inference

In addition to descriptive statistics, the principle method used within the study to estimate WTP is a Bayesian Interval Regression model specified under two different assumptions concerning the error distribution, i.e. the log-Normal and the Normal.

The two models are both of the standard linear form:

$$WTP_i = a + b\ BTSI_i + cZ_i + e_i \quad (1)$$

and

$$\ln(WTP_i) = a + b\ BTSI_i + cZ_i + e_i \quad (2)$$

where

$$BTSI_i = 1\ if\ incentivised\ under\ BSTI\ and\ 0\ otherwise$$
$$Z_i\ is\ a\ vector\ of\ covariates\ describing\ individual\ i$$
$$e_i\ is\ a\ normal\ error\ with\ mean\ 0\ and\ variance\ \sigma^2.$$

The models were estimated with and without $b= 0$ imposed, but we only present the full models below (in Table 4 which also outlines the covariates used) since all other coefficients were virtually the same with and without this restriction. The interval regression assumes that in either case the error is normally distributed with mean zero and constant variance. Interval regression is used to model dependent variables that have interval censoring. That is, each observation is known to lie within a range, but is potentially unobserved within that range. The upper and lower ranges can be bounded or unbounded, and if some of the values of some dependent variables are known exactly, then they can be treated as exact. An option to run interval regressions now exists for many packages including SAS, R, Gretl, and Stata. A full description of the interval model can be found in Stewart (1983). This model is the same whether it is estimated using a Classical or Bayesian approach. The former would map the likelihood function as expressed in Stewart (1983) and the latter can map the posterior (proportional to the likelihood multiplied by a weak prior) using any number of MCMC (Markov chain Monte Carlo) algorithms. The MCMC algorithms used here were coded from first principles in the GAUSS mathematical and statistical system software. Since WTP is only observed as an interval with an upper and lower bound (*l.u*), the WTP or lnWTP is treated as a latent variable with a conditional mean and variance as in the regressions above. Although the Bayesian model was used here with non-informative priors, the results will be virtually identical to those obtained using a Classical approach (Stewart, 1983). The log-Normal model is preferred in this case, given that the distribution of WTP estimates (as averages of the intervals) are clearly of a log-Normal type. However, as will be shown, the key conclusions of the analysis largely remain invariant to this model preference.

The impact of the BTSI on responses was examined in two ways. First, a test was applied to see whether the BTS scores of the incentivised group differed from the non-incentivised group. Then, the WTP distributions were examined to see if they were different across the two populations using an interval regression model with a BTSI dummy variable. The significance of this dummy variable would indicate whether incentivising participants had an impact on respondents' behaviour. In conducting this component normally distributed WTPs and log-Normal WTPs were allowed in order to see whether the results were robust to distributional assumptions.

5. Results

Descriptive statistics

Table 1 contains a summary of descriptive statistics for the sample of 665 responses. The sample contained 55% males and 45% females, with respondents living in households of 2.2 people on average, of whom 0.4 were under 16 years of age. Average household income was £33,500. Fifty-nine percent of respondents reported living in rural settings, i.e. in either villages or provincial towns, with 41% stating that they lived in urban areas. Nearly 16% of respondents reported being members of an organisation concerned with the welfare of animals such as the RSPCA or RSPB[4].

Table 1 –Descriptive statistics for the survey sample

Description	**Value(s)**
Total sample size	665
Gender	Male (55%); Female (45%)
Age distribution (by group)	<40 (26.8%); 40-59 (32.6%) 60+ (40.6%)
Average household size (persons)	2.2
Average number of children in households <16 years	0.4
Average household income (£ p.a.)	33,500

[4] This percentage would appear relatively high but it is likely that respondents interpreted this question quite widely including a range of organisations and relatively lax definitions of what constitutes membership. This percentage was adjusted downwards to 5% when accounting for sampling bias as described below.

Percentage of rural residents	59%
Percent of respondents who are members of an organisation concerned with the welfare of animals	16%
Percentage of respondents who were vegetarians	2.5%
Percentage of respondents who consumed chicken	94.9%
Consumption of meals (from all sources) containing chicken (number of meals / week)	1-2 (51.5%); 3-7 (43.2%); 7+ (5.3%)
Weekly spend on chicken (all sources) (£ /week / household)	0-£4.99 (37%); £5-£14.99 (54.5%); £15+ (8.5%)
Percent of sample making regular purchases of free-range chicken	22.9%

Almost 95% of respondents reported being consumers of chicken. Just under half of those who did not consume chicken stated their reason as being vegetarianism. Respondents consuming chicken reported eating an average of 3.05 meals per week containing chicken (mode 1-2; includes takeaways and meals out), with a range from one to 15. Respondents reported spending a modal value of £5-£10 per week (mean £7.8) on chicken purchases for their household, excluding takeaway meals and meals eaten outside the home (respondents were asked to select from categories for their response, which ranged from zero to more than £20 per week). Chicken from conventional production systems was by far the most common purchase, with free-range chicken regularly purchased by just 23% of the sample. Fifty-six percent of the sample reported purchasing free-range chicken meat only occasionally and 21% never purchased it.

Respondents were asked to rank their concern for farmed animals against a number of other widely held concerns, such as those in relation to the environment. Chicken eaters are most concerned about food safety, with animal welfare concerns near to last. Non-chicken eaters place food safety concerns last, but, again, animal welfare concerns are low in the order of priorities ('healthy diet' was ranked first). Respondents were asked to rank their level of concern for the welfare of broiler chickens reared in the UK on a scale of 0-10, where 0 = not concerned at all and 10 = very concerned. The sample average rank was 8.2, with chicken eaters scoring insignificantly higher

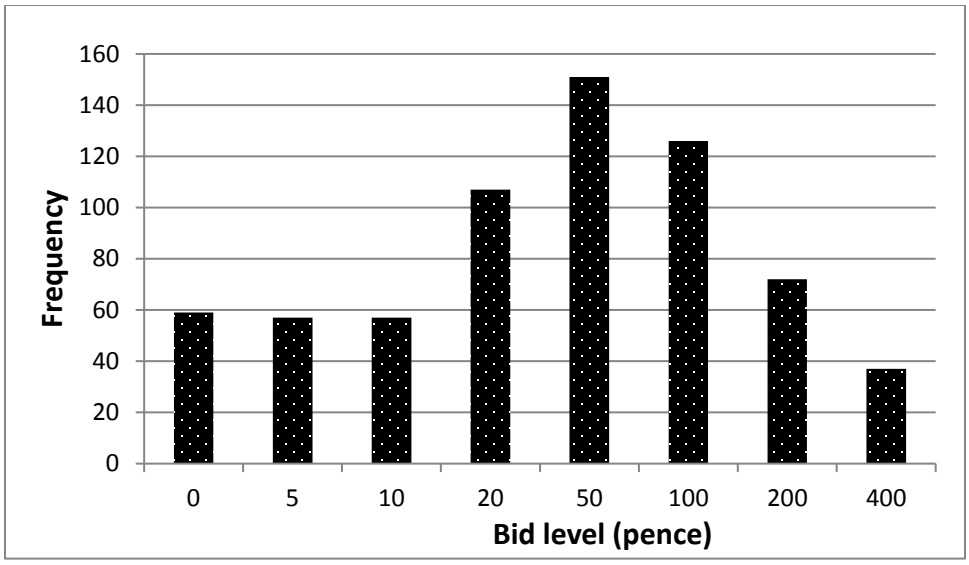
than non-chicken eaters, i.e. 8.6 compared to 8.2. There were significant age differences in ranking of concern, with older respondents ranking concern more highly (F=3.53, p=0.0147).

WTP elicitation

Respondents were asked to indicate the maximum sum that they would be willing to pay per week as an additional amount on the cost of their chicken meat purchases (or, in the case of non-chicken eaters, as an increase in taxation) to support the broiler legislation. Figure 1 shows the distribution of WTP responses. Those classified as zero in Figure 1 are those who responded 'No' to the lowest WTP amount offered to them (5p per week extra) and may therefore include those with a non-zero but low (i.e. less than 5p) WTP. In the estimation of WTP, in the case of the unlogged data, the latent variables for those saying they would not pay the smallest possible amount (5p) are truncated above by the smallest possible values in the survey (5p), and this allows potentially negative WTPs. For the logged version, the logged WTP is truncated above the log of the smallest possible value, except by construction the latent WTPs will (when being anti-logged i.e. taking the exponential of the log WTP) imply positive though negligible WTPs.

Respondents were asked a number of debriefing questions to explore the reasons for their stated WTP, and test the rationality of responses. These questions presented a number of propositional statements and respondents had to express their level of agreement with each, using an 11-point scale where 0 = 'does not reflect my views at all' and 10 = 'reflects my views completely' (see Table 2). Respondents were first asked to indicate whether they understood the information provided in the questionnaire about the broiler legislation. The mean sample rank score was 8.7, indicating a high level of agreement. When asked whether they understood the WTP question (for both payment vehicles), the great majority of respondents indicated that they did (mean score 8.9).

Figure 1. Distribution of sample over bid acceptance levels (pence).



Respondents generally believed that the welfare of chickens reared for meat needed to be improved and also expressed a belief that the Directive itself would act to improve the welfare of chickens reared for meat. It is interesting to note that although the Directive was introduced in 2010 (the study outlined here was part of an ex-post appraisal of the legislation) respondents were generally not aware of the Directive (for example, all focus group participants were found to be unaware of the legislation). There was a relatively high level of agreement by respondents that their WTP values reflected their concerns about the welfare of broiler chickens. Overall, respondents tended to agree with the statement that they should not have to pay more to improve chicken welfare. As might be expected, respondents who did not eat chicken were less likely to agree with the statement. There was no significant difference on this measure between those who regularly purchased free-range chicken and those who did not, even though there was a significant difference in WTP between these groups ($F=13.28$, $p<0.0001$), with regular purchasers willing to pay £0.9/week more than those who never purchased. Perhaps surprisingly, those respondents that most strongly expressed the view that they should not have to pay more (scores of 9 and 10) did not have a significantly lower WTP ($t=-0.04$, $p=0.9680$). Finally, most of the respondents (212/59% scored between 6 and 10) who received the questionnaire version with the BTSI agreed with the statement that they believed the information provided to them about the truth score (12% scored 5 suggesting they neither believed nor did not believe the information and 14% scored between 0 and 4 suggesting they did not believe the information).

Table 2. Responses to WTP debriefing questions, where average rank score expresses degree of agreement with propositional statements on a scale of zero to 10.

Propositional statements	Rank score[a] (0–10)
I understood the information provided to me about the legislation	8.74
I understood the willingness to pay questions	8.94
The welfare of chickens reared for meat needs improving	8.36
The legislation will improve the welfare of chickens reared for meat	8.05
My WTP reflects my concern about chicken welfare	7.57
I should not have to pay more to improve chicken welfare	6.63
I believed the information provided about the Truth Score	6.23

[a] Note: A rank score of 0 = 'does not reflect my views at all' and a rank of 10 = 'reflects my views completely'.

Table 3 shows the numbers of respondents who said that they would or would not pay something for the legislation. It shows that although the majority (63.5%) say they would pay something, 103 (15.5%) individuals reported that they were either definitely not or probably not prepared to pay. Of these, 58 subsequently went on to indicate a non-zero WTP amount when asked how much they would be willing to pay. This could reflect the WTP amounts presented to them, which started at £0.05 per week (0.6% increase on stated chicken expenditure). This may have induced a positive response from respondents who otherwise might have bid nothing. However, given this apparent inconsistency in stated preferences, for those respondents who had previously indicated that they would 'definitely not' or 'probably not' be willing to pay something for the legislation, a zero WTP is assumed in the estimation of WTP below to reduce any upward bias due to 'yea saying'.

Table 3: Numbers of respondents who would be prepared to pay something for the broiler legislation

	Frequency	Percent
Definitely yes (1)	154	23.2

Possibly yes (2)	267	40.3
Not sure (3)	139	20.9
Probably Not (4)	77	11.6
Definitely Not (5)	26	3.9
Total	663	99.9

Regression results

Table 4 gives the interval regression results for both the Normal (in the bottom panel) and log-Normal (in the top panel) models as described in Equations 1 and 2. Income, membership of an organisation concerned with the welfare of animals (labelled RSPCA) and having children under 16 in the household all have a significant influence on WTP in both regressions. The regressors have been demeaned so that for the Normal model the intercept can be interpreted as the mean (and median since it is a symmetric distribution) estimate of WTP at the sample mean of the descriptors. For the log-Normal model, the mean and median WTPs are calculated from the intercept and variance of the log-Normal regression, but may still be interpreted as estimates at the sample mean of the descriptors. The preferred model is the log-Normal, given that the distribution of WTP estimates (as averages of the intervals) are clearly of a log-Normal type.

Table 4. Results from Normal and log-Normal interval regressions.[a,b] Dependent variable is WTP.

	Estimate	**SD Est**	**Pseudo-t**
Log_Normal			
Intercept	3.33	0.05	65.17**
BTSI	0.08	0.10	0.86
ln(Income)	0.29	0.08	3.58**
ln(Age)	-0.01	0.14	-0.07
Female	0.19	0.10	1.82*
RSPCA	0.54	0.14	3.80**
Have Children	0.29	0.13	2.14**
Error Variance	1.65	0.10	
Median WTP	27.97	1.43	
Mean WTP	64.11	4.52	

Normal			
Intercept	56.43	3.07	18.34**
BTSI	4.95	6.13	0.80
ln(Income)	13.20	4.91	2.68**
ln(Age)	-0.71	8.50	-0.08
Female	5.17	6.47	0.80
RSPCA	30.51	8.55	3.56**
Have Children	21.37	8.36	2.55**
Error Variance	5782.59	372.53	
Mean and Median WTP	56.43	3.07	

[a] Asterisks * and ** denote 2-tailed significance at the 10% and 5% significance respectively.
[b] Number of observations = 665

WTP robustness checks

In order to guard against 'yea saying', the models were re-estimated under the assumption that those respondents who had previously indicated that they would 'definitely not' or 'probably not' be willing to pay something for the legislation, had zero or even negative WTPs, even if they subsequently indicated otherwise. In addition, we corrected for bias which may have been caused by over-representation of some classes of socio-demographic stratification in the sample compared with the general population. We did this by inputting average key socio-demographic characteristics of the general population of England and Wales for 2014 (ONS, 2014a) into the preferred model and re-estimating WTP. These key characteristics are income (£27,200), age (40), gender (49.3% male), the proportion of households with children under 16 (30.3%) and an assumption that 5% of the general population were members of an organization concerned with the welfare of animals.

The preferred log-Normal model with these adjustments gives an estimated mean WTP of £62.5 and a median WTP of £21.5 per household per week. Differences between these values reflect the highly diffuse upper tail in the log-Normal distribution. Thus, the median estimate is considered the better measure of central tendency, as a measure of the consumer/citizen surplus

associated with the legislation[5]. These values compare with those of around £64 and £28 respectively of the unadjusted WTP estimates from the log-Normal model shown in Table 4.

However, to explore the robustness of this conclusion further the log-Normal model was re-estimated assigning a zero WTP to 148 respondents who answered 'No' when asked (prior to informing them about the legislation) if they would be willing to pay something to improve the welfare of meat chickens but then subsequently stated a positive WTP to support the broiler legislation. These results are not shown here but are very similar to the log-Normal model results shown in Table 4 except that median WTP was estimated at £14.3 per household per week.

Impact of the BTSI

Before returning to the results in Table 4, it is worth noting that the mean BTS scores for incentivised individuals relative to non-incentivised individuals were 0.17 and 0.13 respectively. While the incentivised group had higher scores on average, a t-test of the difference between group scores was not statistically significant (p-value > 0.15). With respect to the group of individuals with a positive WTP, the incentivised and non-incentivised groups represented 69.5% and 71.2% of these respondents respectively. A z-test for differences in proportions was again statistically insignificant (p value > 0.30).

Next we return to the results in Table 4 to test for the impact of the BTSI on WTPs. As before two forms of interval regression were used, i.e. Normal and log-Normal (see Table 4). The point here is not to make a comparison of the WTP results, but to show that inference is robust to assumptions about the nature of the WTP distributions. As these are Bayesian estimates, the "Pseudo-t" values are the ratio of the estimate divided by the standard deviation of the estimate, similar to the t-statistic in classical regression. If this statistic exceeds 1.96 then it is likely that the associated regressor is having a non-zero impact on the dependent variable. The BTSI dummy indicates whether individuals have been incentivised according to the BTSI.

[5] Note that the very few in the sample (34 of 665) who did not eat chicken had an average WTP of around twice that of chicken eaters.

The coefficient of BTSI can be interpreted as the increase in WTP (directly for the Normal or logged WTP for the log-Normal case) resulting from the BTSI treatment. The coefficients are positive for both the Normal (4.95) and log-Normal (0.09) cases, but given their large standard deviations, these cannot be reliably said to have a positive impact on WTP. In addition to the regression results shown in Table 4, we performed a regression for each distribution with the BTSI dummy as the sole explanatory variable, in order to test that the lack of significance for BTSI is not due to collinearity of the BTSI dummy with other regressors. Again, the coefficents of BTSI were positive but not significant. From these results, it is evident that according to either distribution there is no statistically-significant evidence that the use of the BTSI has had an impact on WTP in this case.

In summary, the results show that the responses of people with respect to the truth score itself, their WTP or the propensity to give inconsistent responses are largely invariant to the BTSI.

6. Discussion

Our survey results indicate that people in England and Wales generally support the broiler Directive legislation and appear to have a substantial WTP to support it. WTP was found to be positively correlated with income, level of concern about chicken welfare and belief that the legislation will improve the welfare of meat chickens, as would be expected. Moreover, a large majority of respondents stated that they understood both the information provided to them about the Directive and the WTP questions they were asked to respond to. Thus, the survey instrument appeared to work well and to result in credible responses from respondents. The sample size of 665 respondents meets the requirement recommended by Mitchell and Carson (1989 p. 229) for a sample size of 600 usable responses or more for estimating benefits for policy purposes.

However, there are many opportunities for various types of bias within stated preference survey instruments and there is a vast literature on such biases within stated preference research, including the issue of misreporting. In the first place, since the legislation already exists, fully informed and economically rational participants would not be expected to be willing to pay anything more for what they already have and are already paying for, one way or another. However, they may be willing to pay an additional amount to retain the legislation rather than it

being removed (i.e. up to the point where their consumer surplus is zero). Harvey and Hubbard (2013) provide a critique of the elicitation of people's WTP in relation to animal welfare together with a more general thought-provoking consideration of the application of economics to animal welfare. Emotive issues such as animal welfare are susceptible to social desirability bias, where respondents feel pressured to provide a socially desirable response (i.e. to show concern for the humane treatment of animals), which may have been provoked by some questions in our survey. We also acknowledge the possibility of starting point bias and/or range bias associated with the WTP values presented to participants (where they may feel that either the first amount or the range of amounts presented to them is some indication of the amount they should be willing to pay). In addition, 'strategic bias' may have been present, especially in relation to the scenario presented to respondents which described the legislation as being currently evaluated (which was true). Respondents may have felt that implementation of the legislation was under threat and may therefore have strategically over-stated their WTP to ensure that the legislation was implemented. We have used a relatively cautious approach to WTP estimation to guard against these potential biases.

Otherwise, responses of individuals may be shaped (perhaps resulting in misreporting) by the way in which questions are presented (Kahneman and Tversky, 1984), the information they are presented with (Ajzen *et al*., 1996; Hensher, 2006), hypothetical bias (the difference between true WTP and stated WTP) including the perceived social desirability associated with questions (Nederhof, 1985) such as humane treatment of chickens in this survey, the time they are given to complete the survey (Conte *et al*., 2016), the length of the survey (Savage and Waldman, 2008) and so on. It is not possible to test for all possible biases, priming or other effects of survey design. Our survey instrument was carefully designed, pre-tested and piloted to minimize bias in responses and the results analysed and presented to avoid estimation bias. Responses were scrutinised in relation to their credibility, rationality and consistency to identify potential issues that might cast doubt on the validity of the survey instrument design, people's responses to the questions and the subsequent results. Questions regarding respondents' consumption of chicken and their attitudes toward animal and chicken welfare were placed prior to information on the Directive and elicitation of WTP to encourage respondents to think about the importance of chicken in their diets

and the importance that they give to animal and chicken welfare (e.g. in relation to other concerns such as food safety and a healthy diet).

Loomis (2014) identifies and reviews both *ex ante* and *ex post* strategies for overcoming hypothetical bias in stated preference surveys. The former includes 'consequentiality designs' which means that the survey has some potential effect on the utility of respondents such as higher prices or taxes. This approach was used in the survey reported here together with binary, dichotomous choice question formats and compulsory payment mechanisms as recommended by Loomis (see also Carson and Groves, 2007). 'Honesty and realism' approaches are also recommended such as the one used in this survey which involved a request to respondents to "consider your answers to questions carefully, answer honestly and take care to avoid mistakes". In addition, the BTSI was used to further encourage truthful responses. To reduce the possibility of hypothetical bias respondents were also reminded of their budget constraints. Loomis suggests that social desirability bias and cognitive dissonance (where respondents gain utility by responding according to perceived social norms rather than their own personal values) are reduced by having multiple bid values (with small positive amounts) as used in this survey, and by making responses impersonal and anonymous (again as used in the internet survey of this study). In addition, asking respondents what they think others would pay is a way of potentially gauging over-statement of WTP due to social desirability bias (see Lusk and Norwood, 2009) – an approach also used in this study and used as an integral part of the BTS. Ex post methods to reduce hypothetical bias identified by Loomis include reporting median WTP to minimize the effect of implausibly high WTPs, recording respondent uncertainty in relation to their WTP responses and recoding of 'Yes' responses as 'No' where there is sufficient uncertainty regarding respondents having a true positive WTP, all of which were used in this study. It is also noted that using a private good (such as chicken meat in this study) reduces hypothetical bias. Loomis warns that there is no consensus regarding the best method to correct for hypothetical bias, that measuring hypothetical bias is difficult (the analyst needs to know the true WTP) and that it is possible to over-correct for this bias and so underestimate WTP.

We estimate the median WTP as a measure of the consumer surplus associated with the broiler legislation, as £21.5/household/year from the preferred log-Normal Bayesian Interval Regression

model, after correcting for bias, representing 5% of the total annual amount that respondents estimated they spent on chicken meat. Our more conservative estimate, assuming a zero WTP for those who said they would not be willing to pay more, despite subsequently providing a positive response, is £14.3 per household per week.

According to the UK National Census (ONS, 2014a), there were 23.4 million households in England and Wales in 2011. Using the WTP estimate of £21.5/year from above, the aggregate amount that consumers in England and Wales would be willing to pay for the broiler Directive is £504 million/year. The more conservative estimate generates aggregate benefits of £333 million/ year. A 2014 survey of 119 commercial conventional broiler producers estimated the cost of compliance with the broiler Directive in England and Wales over the period 2010 to 2014 to be £108.4 million (current prices), an average of approximately £21.7 million per year or nearly £21.9 million per year including government inspection and enforcement costs (Defra, 2017; Defra 2010), implying a benefit-cost ratio of 23:1. The more conservative estimate gives a benefit-cost ratio of 15:1. Some radical changes in assumptions for both benefits and costs would be required to conclude that the costs of the legislation exceed the benefits.

Prior to the introduction of the Directive, SAC Commercial Ltd (SAC, 2005) estimated that consumers in England and Wales would be willing to pay the equivalent of just under £20/household/year (2014 prices) which is similar to the estimate generated above.[6] However, people's perceptions of the value of the legislation may have been somewhat different before and after implementation of the legislation so the two estimates are not wholly comparable.

With regard to the BTS, we found no statistically significant impact of the BTSI on responses to the survey. We believe this negative finding is at least as important as those studies which suggest that the BTS has an impact. While we would encourage further investigation of this approach, we believe that the literature should report cases where this technique has not demonstrably changed behaviour in addition to occasions when it has. Not to do so would lead to

[6] The SAC estimate of £7.53/person/year. According to the National Census (ONS, 2014a), average household size in 2013 was 2.37 persons with 1.86 adults per household. Therefore, the SAC (2005) WTP estimate equates to £14/household/year, in 2005 prices. Adjusting for inflation (ONS, 2014b), the SAC estimate is just under £20/household/year.

a biased representation of the weight of evidence in favour of positive BTS effects (more generally, for the importance of reporting negative findings we refer readers to the debate concerning 'priming' (Shanks *et al*. 2013)).

There are a number of reasons why our study may have failed to observe a significant BTSI impact. First, while there may have been misreporting by the sample, the level or nature of the incentive may not have been sufficient to alter respondent behaviour. More specifically, the incentive provided for truth telling (the 'prize draw') may have been too small and uncertain, with respondents perceiving only a small chance of gaining financially for 'truth telling' (with an unknown probability of winning because respondents did not know the number of people taking part in the prize draw). Second, although most respondents had some level of belief in the truth score there was an element of disbelief which may have reduced the impact of the BTSI across the sample. Third, irrespective of the efficacy of the incentive, there may simply have been no significant misreporting in the sample and therefore no error to correct. This effect may be the case because the perceived gravity of the issue being investigated encourages truthful reporting. It is worth noting that some previous studies which have found the BTS to be effective asked respondents to report on what might be deemed more trivial matters, for example surveys of the extent of people's recognition of world leaders' names and film titles (see Weaver and Prelec, 2013). Arguably, the more gravity that respondents attach to the issue under investigation, the more likely respondents are to carefully consider the questions presented to them, thus reducing one potential source of misreporting. Additionally, a number of aspects of the survey and questionnaire were designed to minimise bias and reduce any potential misreporting and this may have contributed to an absence of misreporting by respondents.

7. Conclusion

We find that people in England and Wales have a substantial stated WTP to support legislation to monitor and improve the welfare of broiler chickens in the UK. At a societal level, the benefits of the legislation, as measured by people's estimated WTP, greatly outweighed the costs (as estimated by a survey of broiler producers). This suggests both that the current broiler legislation was worth implementing and that it is worth continuing in the UK. Use of a Bayesian Truth Serum Incentive (BTSI) method to encourage honest and careful responses made no statistically

significant difference in terms of people's responses or their WTP. This finding suggests that further studies need to be undertaken to better determine the exact circumstances under which the use of the BTSI has most effect in reducing misreporting.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Appendix S1 Specimen copy of the online questionnaire

Appendix S2 Specification of the Bayesian Truth Serum scoring formula

References

Ajzen, I., Brown, T.C. and Rosenthal, L.H. 'Information bias in contingent valuation: effects of personal relevance, quality of information and motivational orientation', *Journal of Environmental Economics and Management*, Vol. 30, (1996) pp. 43-57.

Bateman, I.J., Carson, R.T., Day, B., Hanemann, M., Hanley, N., Hett, T., Jones-Lee, M., Loomes, G., Mourato, S., Ozedemiroglu, E., Pearce, D., Sugden, J. and Swanson, J. *Economic valuation with stated preference techniques: A manual* (Cheltenham, UK: Edward Elgar, 2002).

Bennett, R., Kehlbacher, A. and Balcombe, K. 'A method for the economic valuation of animal welfare benefits using a single welfare score', *Animal Welfare*, Vol. 21(S1), (2012) pp. 125-130.

Bennett, R. and Appleby, M. *Animal welfare policy in the European Union.* In: Oskam, A., Meester, G. and Silvis, H. (eds.) EU policy for agriculture, food and rural areas. 2nd edition. Wageningen Academic Publishers, (2011) pp. 249-258.

Bennett, R., Butterworth, A., Jones, P., Kehlbacher, A. and Tranter, R. *Valuation of animal welfare benefits*. Report to the Department for the Environment, Food and Rural Affairs (University of Reading: Reading, UK, 2011).

621 Carson, R.T. and Groves, T. 'Incentive and informational properties of preference questions',
622 *Environmental and Resource Economics*, Vol. 37(1), (2007) pp. 181-210.
623
624 Clark, B., Stewart, G.B., Panzone, L.A., Kyriazakis, I. and Frewer, L.J. 'A systematic review of
625 public attitudes, perceptions and behaviours towards production diseases associated with farm
626 animal welfare', *Journal of Agricultural and Environmental Ethics*, Vol. 29(3), (2016) pp. 455-
627 478.
628
629 Conte, A., Scarsini, M. and Sürücü, O. 'The impact of time limitation: insights from a queueing
630 experiment', *Judgment and Decision Making*, Vol. 11(3), (2016) pp. 260-274.
631
632 Defra. *Study to evaluate the effectiveness of Regulation (Directive 2007/43/EC) in England and*
633 *Wales*. EVID4 Evidence Project Final Report (Rev. 06/11) (2017).
634
635 Defra Council Directive 2007/43/EC. *Laying down minimum rules for the production of chickens*
636 *kept for meat*, Impact Assessment No. Defra1000 (Defra: London, 2010).
637
638 European Commission Council Directive 2007/43/EC. *Laying down minimum rules for the*
639 *protection of chickens kept for meat production*. Official Journal of the European Union, L 182/19,
640 12.7.2007. (2007).
641 http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32007L0043&rid=4
642
643 European Commission, Animal Welfare. (2016)
644 http://ec.europa.eu/food/animals/welfare/index_en.htm
645
646 Harvey, D. and Hubbard, C. 'Reconsidering the political economy of farm animal welfare: An
647 anatomy of market failure', *Food Policy*, Vol. 38, (2013) pp. 105-114.
648
649 Heinzen, R.R and Bridges, J.F. 'Comparison of four contingent valuation methods to estimate
650 the economic value of a pneumococcal vaccine in Bangladesh', *International Journal of*
651 *Technology Assessment in Health Care*, Vol. 24(4), (2008) pp. 481-487.

Hensher, D.A. 'How do respondents process stated choice experiments? Attribute consideration under varying information load', *Journal of Applied Econometrics*, Vol. 21, (2006) pp. 861-878.

Hurwicz, L. 'The design of mechanisms for resource allocation', *American Economic Review*, Vol. 63, (1972) pp. 1-30.

John, L.K., Loewenstein, G. and Prelec, D. 'Measuring the prevalence of questionable research practices with incentives for truth telling', *Psychological Science*, Vol. 23(5), (2012) pp. 524-532.

Kahneman, D. and Tversky, A. 'Choices, values, and frames', *American Psychologist*, Vol. 39, (1984) pp. 341–350.

Lagerkvist, C.J. and Hess, S. 'A meta-analysis of consumer willingness to pay for farm animal welfare', *European Review of Agricultural Economics*, Vol. 38(1), (2011) pp. 55-78.

Loomis, J.B. 'Strategies for overcoming hypothetical bias in stated preference surveys', *Journal of Agricultural and Resource Economics* Vol 39(1), (2014) pp. 34-46.

Lusk, J.L. and Norwood, F.B. 'An inferred valuation method', *Land Economics* Vol 85, (2009) pp. 500-514.

Mayfield, L.E., Bennett, R.M., Tranter, R.B. and Wooldridge, M. J. '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*, *Food and Agriculture*, Vol. 15(3), (2007) pp. 59-73.

McFadden, D. Contingent valuation and social choice. American Journal of Agricultural Economics Vol 76, (1994) pp. 689-708.

682 Mitchell, R.C. and Carson, R.T. *Using surveys to value public goods. The Contingent Valuation*
683 *Method.* (Resources for the Future: Washington D.C., 1989)

684 Nederhof, A.J. 'Methods of coping with social desirability bias: a review', *European Journal of*
685 *Social Psychology*, Vol 15, (1985) pp. 263-280.
686
687 ONS, *Families and Households, 2013*. (2014a)
688 http://www.ons.gov.uk/ons/publications/re-reference-tables.html?edition=tcm%3A77-328237
689
690 ONS, *Consumer Price Inflation*, September 2014. (2014b)
691 http://www.ons.gov.uk/ons/publications/re-reference-tables.html?edition=tcm%3A77-323601
692
693 Pearce, D. and Ozdemiroglu, E. *Economic valuation with stated preference techniques*,
694 (Department for Transport, Local Government and the Regions: Rotherham, 2002).
695
696 Prelec, D. 'A Bayesian truth serum for subjective data', *Science*, Vol. 306, (2004) pp. 462-466.
697
698 SAC, *Estimating non-market benefits of reduced stocking density and other welfare increasing*
699 *measures for meat chickens in England*, Report to Department of Food and Rural Affairs: Project
700 Code AW0236, 2005.
701
702 Savage, S.J. and Waldman, D.M. 'Learning and fatigue during choice experiments: a comparison
703 of online and mail survey modes', *Journal of Applied Econometrics*, Vol. 23(3), (2008) pp. 351-
704 371.

705 Shanks, D.R., Newell, B.R., Lee, E.H., Balakrishnan, D., Ekelund, L., Cenac, Z., Kavvadia, F.
706 and Moore, C. 'Priming intelligent behavior: an elusive phenomenon', *PLoS ONE*, Vol 8(4),
707 (2013), e56515.
708
709 Stewart, M. B. On least squares estimation when the dependent variable is grouped. *Review of*
710 *Economic Studies* Vol 50, (1983) pp. 737-753.
711

Weaver, R. and Prelec, D. 'Creating truth-telling incentives with the Bayesian Truth Serum', *Journal of Marketing Research*, Vol. 50(3), (2013) pp. 289-302.

Zhao, J and Kling, C.L. Willingness to pay, compensating variation and the cost of commitment. *Economic Inquiry* Vol 42(3), (2004) pp. 503-517.

Appendix 2: BTS (Bayesian Truth Serum) scoring formula

If question answers and predictions are denoted by

$$x^r = (x_1^r, .., x_m^r)(x_k^r \in \{0,1\}, \sum_k x_k^r = 1) \quad (1)$$ and

$$y^r = (y_1^r, .. y_m^r)(y_k^r \geq 0, \sum_k y_k^r = 1) \quad (2)$$ respectively,

where x^r are the answers of respondent, *r* to questions 1 to *m* and y^r are the predictions of respondent, r to questions 1 to *m*, then we can calculate the population endorsement frequencies, $\bar{x}_k$, and the (geometric) average, $\bar{y}_k$, of predicted frequencies as:

$$\bar{x}_k = \lim_{n\to\infty} \frac{1}{n} \sum_{r=1}^{n} x_k^r, \quad (3)$$

$$\log \bar{y}_k = \lim_{n\to\infty} \frac{1}{n} \sum_{r=1}^{n} \log y_k^r \quad (4)$$

where *n* is the sample size. Instead of applying a preset answer key, we evaluate answers according to their information score, which is the log-ratio of actual-to-predicted endorsement frequencies. The information score for answer *k* is

$$\log \frac{\bar{x}_k}{\bar{y}_k} \quad (5)$$

The total score for a respondent combines the information score with a separate score for the accuracy of predictions:

$$\text{score for respondent } r = \text{information score} + \text{prediction score} =$$

$$\sum_k x_k^r \log \frac{\bar{x}_k}{\bar{y}_k} + \alpha \sum_k \bar{x}_k \log \frac{y_k^r}{\bar{x}_k}, 0 < \alpha \quad (6)$$

749
750
751 (Prelec, 2004).
752
753
754