Journal of Agricultural and Resource Economics 44(2):345–361 ISSN 1068-5502

Is There a "Cage-Free" Lunch in U.S. Egg Production? Public Views of Laying-Hen Housing Attributes

Daniel Ochs, Christopher A. Wolf, Nicole J. Widmar, and Courtney Bir

Animal welfare–related production attributes are increasingly considered by U.S. consumers making food purchasing decisions and U.S. voters at the ballot box. This research considers U.S. consumer preferences for egg production attributes. The results reveal preferences for less hen stress, more natural hen behavior, and improved worker health and welfare. We propose an index combining animal welfare scores and consumer preference shares for determining preferred combinations of egg production attributes. When weighting hen housing systems by consumer preference for animal and worker welfare attributes, the preferred system is enriched colony housing, which differs from recent retailer commitments to cage-free aviaries.

Key words: animal welfare index, best–worst scaling, egg industry, hen welfare, preference elicitation

Introduction

For many years, the cage confinement systems used by U.S. egg industry have not allowed for many natural hen behaviors (Watnick, 2016). Recently, the U.S. public has mandated changes to laying-hen housing systems away from these "conventional cages" through legal and market channels. Legal changes have occurred in many states in the form of legislation (e.g., Michigan PA-0117 and Florida Amendment 10) and ballot initiatives (e.g., California's Prevention of Farm Animal Cruelty Act [Proposition 2]). Both market and legal changes mandate minimum space and conditions that allow other natural hen behaviors (e.g., perching, scratching, and spreading of wings). The legal mandates do not specify the housing system, and multiple candidates meet the requirements, including enriched colony housing and cage-free aviaries (Matthews and Sumner, 2015).

Food retailers, including grocers and restaurants, have been increasing egg supplies sourced from operations that do not use conventional battery cage systems, with a focus on cage-free aviary production. For example, with purchases of near 2 billion eggs annually in the United States, McDonald's has committed to moving away from using eggs produced from hens raised in cages for its restaurants in the United States and Canada over the next decade (Jargon and Beilfuss, 2015). Walmart, Inc., the largest grocer in the United States, has committed to do the same by 2025 in both its Walmart and Sam's Club retail locations (Walmart, 2016).

The candidate hen housing systems that are alternatives to conventional cages present trade-offs across hen welfare, worker welfare, environmental impact, cost, and productivity attributes. Lay et al. (2011) found no single superior housing system and concluded that the right combination of housing design, breed, rearing conditions, and management was essential to optimize hen welfare

Daniel Ochs is an extension educator and Christopher A. Wolf (corresponding author) is a professor in the department of Agricultural, Food, and Resource Economics at Michigan State University. Nicole J. Widmar is a professor and Courtney Bir is a graduate research assistant in the Department of Agricultural Economics at Purdue University.
The authors gratefully acknowledge funding from the Henry and Annabel Larzelere graduate fellowship.
Review coordinated by Darren Hudson.

and productivity. Blatchford, Fulton, and Mench (2016) documented trade-offs between hen housing systems using a long-term hen housing study conducted by the Coalition for Sustainable Egg Supply (Coalition for Sustainable Egg Supply, 2015), which found that, for example, cannibalistic behavior and microbial levels are greater in enriched colony housing and cage-free aviary systems compared to conventional cage systems. Matthews and Sumner (2015) also noted that eggs laid outside the nest box in cage-free aviary systems lead to more uncollectable, downgraded, or unmarketable eggs.

With respect to public or consumer preferences, hen housing systems can be thought of as a set of attributes. For example, the systems may differ with respect to hen stress and behavior, cannibalism and aggression, worker health and safety, and environmental impact. No housing system is superior across all dimensions of hen welfare. There are also trade-offs in system attributes with respect to worker welfare and safety and environmental impacts. The sum of those attributes can define system welfare from a public or consumer perspective. Given the shift toward alternative hen housing systems, it is important to better understand specific welfare trade-offs between systems in conjunction with consumer preferences to ensure that industry housing investment aligns with true consumer preference for welfare attributes.

Animal welfare and consumer research methods to elicit and understand consumer preferences and values include Likert scales, forced ranking, and best–worst scaling. This research uses a nationally representative sample of respondents to an online survey to examine U.S. public preference for hen housing system attributes. Using multiple methods, we assess the robustness of public opinion and consistency across methods. Further, utilizing the preference shares from best–worst analysis, we develop a welfare index as a composite of housing system attributes.

Egg Production Housing System Attributes

Swanson et al. (2011) discussed the challenge presented by laying-hen welfare and concluded that the results depended on the assessment perspective. Yilmaz Dikmen et al. (2016) agreed and suggested that trade-offs exist between hen housing systems, with no single system claiming superior design. Bringing together a diverse group of animal welfare and industry experts, the Coalition for Sustainable Egg Supply (2015) documented the strengths and weaknesses of hen housing systems in a long-term study. Following CSES, the relevant commercial hen housing systems may be described as

1. Conventional cage housing: Cages housing 4–9 hens. Approximate space per hen is 80 square inches. There may be thousands of cages, often totaling 200,000 hens in each building.

2. Enriched colony housing: Enriched colony systems house 60–250 hens in cages that are larger than conventional cages and equipped with perches, nesting areas, and material to facilitate foraging and dust bathing. Approximate space per hen is 116 square inches. These systems can house 100,000 or more hens in each building.

3. Cage-free aviary housing: Cage-free aviary systems allow hens to roam throughout various-sized sections of a building. Each section contains perches, nesting areas, and dust-bathing material. Approximate space per hen is 144 square inches. There can be 80,000 or more hens per building.

The full list of attributes examined by CSES is comprehensive and technical. To build on their work and connect welfare assessment to the U.S. consumer, this research employs CSES attribute sets that accurately represent welfare considerations for hen housing systems. The attributes utilized were those that CSES used to differentiate and describe the welfare aspects of the housing system coupled with aspects that have received media attention (Mayer, 2015; Bruilliard, 2017; Whoriskey, 2017). The set of attributes studied here included hen mortality, hen behaviors (i.e., scratching, perching, dust bathing), hen cannibalism/aggression, hen stress physiology, hen feather condition, hen foot condition, and worker health and safety. Throughout this research, these attributes shall be referred to as worker health and safety, mortality, behavior, cannibalism/aggression, stress, feather condition, and foot condition. The following descriptions elaborate on the attribute implications and explore some of the baseline results.

Mortality

The main causes of death in all three housing systems were hypocalcemia and egg yolk peritonitis (Coalition for Sustainable Egg Supply, 2015). Other causes of death were highest in the cage-free aviary and included being caught in the structure, excessive pecking, and emaciation. Conventional cages and enriched colonies housing systems were similar across these other causes.

Foot Condition

Foot condition was scored using the Welfare Quality Assessment protocol for poultry (Netherlands Welfare Quality® Consortium, 2009). This scoring method looked at foot details such as toe damage and claw length. Foot issues were more common in the conventional cages but tended to be more severe in the cage-free aviary, with intermediary results in the enriched colony.

Hen Stress

Short-term stress levels were measured with blood samples to look at heterohpil-to-lymphocyte ratios and white blood cell counts at pullet placement and at several stages during the laying cycle (Coalition for Sustainable Egg Supply, 2015). Additionally, the weights of adrenal glands were taken for hens sampled at the end of the laying period as a measure of long-term stress. No significant differences in hen stress levels were found among the three housing systems.

Hen Behavior

Behavior refers to the utilization of the housing space and enhancements such as scratch pads, perches, and nests. Nest usage was scored by counting the number of eggs laid inside versus outside of the nests. As might be expected, laying hens were able to exhibit more natural behaviors in the cage-free aviary, slightly less in the enriched colony, and the least in the conventional cage system.

Cannibalism/Aggression

CSES researchers examined the number of deaths caused by excessive pecking to gauge the level of cannibalism and aggression occurring in any given housing system. Higher levels were documented in the more open cage-free aviary system.

Feather Condition

Also utilizing the Welfare Quality Assessment protocol for poultry (Netherlands Welfare Quality® Consortium, 2009), feather condition was scored by looking at hen feather loss and feather cleanliness. Feathers were dirtiest in the cage-free aviary, but there was greater feather loss in the conventional cages, with intermediary results in the enriched colony.

Worker Health and Safety

CSES addressed this category by looking at several subcategories, including worker exposure to particulate matter, ammonia, and endotoxin; worker lung health, worker ergonomics, worker access,

Table 1. CSES Differences from Conventional Cage Baseline

Welfare Attribute	Enriched Colony	Cage-Free Aviary
Hen mortality	0	−2
Hen foot condition	1	0
Hen behavior	2	3
Hen stress	0	0
Hen cannibalism/aggression	−1	−3
Hen feather condition	0	1
Worker health and safety[a]	0	−1

Notes: Coalition for a Sustainable Egg Supply (CSES) experts used research to score each attribute as change from conventional cage systems, in which negative values were negative impacts (−4 = exceptionally worse, −3 = substantially worse, −2 = worse, −1 = slightly worse), 0 was similar, and positive values were positive impacts (+1 = slightly better, +2 = better, +3 = substantially better, +4 = exceptionally better).
[a] Adapted by authors from the Coalition for Sustainable Egg Supply (2015) by averaging across worker health and safety subcategories to achieve precisely −1.14, but rounded to −1 for inclusion in the table.
Source: Coalition for Sustainable Egg Supply (2015).

and worker egress. Breathing exposure was measured using exposure monitors, and lung health was measured at the beginning and end of a worker's shift by determining pulmonary function and respiratory symptoms. Movement/ergonomics was scored by classifying tasks performed in each system into different risk categories. Due to the open nature of the cage-free aviary system, more eggs were laid outside of nesting areas, which made egg collection more strenuous for workers and created higher levels of airborne particulates. Worker health and safety levels were the same in conventional cage and enriched colony housing.

The CSES measurements used a rating scale from +4 to −4 for each welfare attribute (Table 1).[1] The conventional cage housing system was used as a baseline against which other housing systems were compared. A score of +4 indicated that the housing type had an "exceptionally better" effect on the attribute in question. Conversely, a score of −4 indicated that the housing type performed "exceptionally worse" with respect to that attribute. A score of 0 indicated no difference between the conventional cage and the housing type in question. All the attributes discussed above and summarized in Table 1 were directly measured on this scale except for worker health and safety. To create a single score for this category from the CSES documentation, the worker health and safety subcategories were averaged.

Methods

Survey

A national online survey was administered in April 2017 to collect information about U.S. public perceptions of hen housing types and laying-hen welfare. The survey was written by a team of researchers from Michigan State University and Purdue University. The survey was anonymous and approved by both Michigan State University and Purdue University Human Research Protection Programs. The survey was coded using the Qualtrics online platform and administered by Lightspeed GMI to generate a nationally representative U.S. sample in terms of gender, age, income, region of residence, and education level.

Respondents were asked about the impacts they expected a shift from conventional housing to the other housing types would have on a series of welfare attributes. Following this series of

[1] CSES experts utilized research to score each attribute as change from conventional cage systems, where negative values were negative impacts (−4 = exceptionally worse, −3 = substantially worse, −2 = worse, −1 = slightly worse), 0 was similar, and positive values were positive impacts (+1 = slightly better, +2 = better, +3 = substantially better, +4 = exceptionally better).

questions about expected impacts of a housing type, respondents were asked to prioritize hen housing system attributes studied by CSES using several preference elicitation methods. This allowed us to determine the preferred housing type using a portfolio of attributes.[2]

Likert Scaling

First published by Renis Likert in 1932, the Likert or monadic rating scale method of eliciting preference or perception is widely used across disciplines (Willits, Theodori, and Luloff, 2016). Their extensive use is due in large part to their ease and ability to evaluate numerous statements efficiently in a survey. In fact, Likert scales have become so common that many researchers do not consider utilizing other methods (Jaeger et al., 2008). Despite the defense of the Likert scale by Willits, Theodori, and Luloff, there is growing support for alternative methods. Cohen (2003), Cohen and Orme (2004), and Hein et al. (2008) all found that the best–worst scaling method has a superior ability to differentiate between items or attributes (i.e., more pairs of samples or attributes were found to significantly differ from one another). Because Likert ratings do not force trade-offs, it is possible to score multiple attributes at the same level, which does not allow for direct comparison or differentiation. Finn and Louviere (1992) cited this lack of discrimination as a primary motivation to develop the best–worst scaling method. For this survey, each of the seven welfare attributes was rated by survey participants on a Likert scale from one ("not important") to five ("extremely important"), with an "I don't know" option to differentiate from neutral choices. To compare the relative importance of attributes using the Likert score, respondents who answered "I don't know" for any attribute were dropped from the sample for all Likert analysis. For example, a respondent who answered "I don't know" for the feather condition attribute was also dropped from the estimation of mean values for all other welfare attributes. In total, this resulted in dropping 894 observations. Although this resulted in a smaller sample size for the Likert analysis, it also eliminated a neutralizing effect from uncertain respondents. Despite this, the Likert method still did not statistically differentiate the relative importance for five of the seven welfare attributes.

Forced Rank

Forced rank is a "single multinomial choice in which respondents are asked not just for their most preferred alternative but also a complete ranking exercise over all alternatives presented" (Carson and Louviere, 2011, p. 548). Although this method has a long history and hypothetically forces the respondent to compare all subsets of attributes, the difficulty of the task grows significantly as the number of attributes increases (Carson and Louviere, 2011). The forced rank question in the survey was straightforward. Respondents were instructed to order the hen and worker welfare attributes, assigning each a value of 1 through 7, indicating most to least important. To complete the task, survey respondents "dragged and dropped" welfare attributes into the order of most to least important.

Best–Worst Scaling

First introduced by Finn and Louviere (1992), best–worst scaling is now used for preference elicitation across many disciplines (Jaeger et al., 2008). A strength of best–worst models is the direct trade-off that respondents are forced to make among a reduced list of attributes (Lusk and Briggeman, 2009; Flynn et al., 2007; Bir, Widmar, and Croney, 2017). With other methods—such as Likert scales—there is no forced decision, potentially resulting in uniform rating for all attributes

[2] Respondents were not supplied with information about the housing systems provided other than the name, as might appear on a label. These results mimic actual market behavior given their existing set of knowledge, opinions and perceptions about hen housing systems. We might reasonably expect most people to have some familiarity with some hen housing systems such as "cage-free" or similar labeling since it is common on egg cartons across the United States.

(Jaeger et al., 2008). Similarly, ranking imposes trade-offs but assumes that the respondent has adequately compared every subset in the format imposed by the best–worst choice design. Marley and Louviere (2005) suggested that the best–worst task was easier for survey participants than forced ranking. This result was supported by Jaeger et al. (2008), who found that 63% of participants found the best–worst task easier than forced ranking. Cohen (2003) suggested that part of these results may be due to the greater cognitive effort being put forth and nearly double the time required to complete the task. Given the significant support for best–worst as a superior method of preference elicitation, a ranking of individual-level preference shares may be more valuable than either forced ranking or Likert scaling. Louviere (2004) suggested that binary choices are more natural to humans. Based on this human predisposition to more easily select extreme values, it is logical to expect a best–worst measurement to have superior performance over a simple ranking task. Hein et al. (2008) found that the best–worst scaling method allowed for greater discernment in food choices than either unstructured scale ranks or forced preference rank, further supporting the strengths of this method.

To assess the welfare attributes deemed most and least important, survey respondents were shown seven choice sets made up of unique combinations of four hen and worker welfare attributes. In each of the seven distinct combinations of four factors, respondents indicated the attribute they considered to be most important and the attribute they considered to be least important. The design of the seven choice sets with four choice combinations was developed to maximize D-efficiency using an orthogonal experimental design. Each of the seven welfare attributes could have been chosen four times. Intuitively, the pair of welfare attributes chosen as most important and least important for any given choice set represents a maximum difference in perceived importance between those two attributes for the respondent. If the seven attributes are thought of as continuous—from most important to least important—from the perspective of any given respondent with a particular position λ_j along a continuous scale of relative importance, then a maximum likelihood estimate (MLE) can be used to estimate that particular location (Lusk and Briggeman, 2009). The relative importance of a welfare attribute to a respondent i can be calculated as the MLE position plus or minus some given level of random error, ε_{ij}:

$$I_{ij} = \lambda_j + \varepsilon_{ij} \tag{1}$$

The probability that respondent i chooses welfare attribute j as "most important" and welfare attribute k as "least important" can be thought of as the probability that attributes j and k are further apart than any other combination of presented attributes along the continuous scale. In other words, the absolute value of $I_{ij}I_{ik}$ is greater than the difference between any other combination of the four welfare attributes presented in a single set.

Under these assumptions, the probability of choosing a particular set of attributes as most important and least important follows a multinomial logit model (MNL) (Lusk and Briggeman, 2009):

$$Prob(j = most \cap k = least) = \frac{e^{\lambda_j - \lambda_k}}{\sum_{l-1}^{J}\sum_{m=1}^{J} e^{\lambda_l - \lambda_m} - J}. \tag{2}$$

Following Lusk and Briggeman (2009), we estimated a random parameters logit (RPL) to allow for heterogeneity (Appendix A). The MNL model alone assumes homogeneity of respondents, which is inherently flawed given our targeting of a representative population of the current U.S. public. Models were estimated using NLogit 5.0.

Following Wolf and Tonsor (2013), we calculated preference shares to examine the relative importance of welfare attributes in terms of the proportion of total preference that consumers place on each welfare attribute. This estimation of weights allows us to assess perceptions of the attributes themselves and facilitates comparison to the other preference elicitation methods. When converting to preference shares, the summed preference of all seven welfare attributes must be equal to 1, and the probability that a particular factor is selected as the most important is

$$share_j = \frac{e^{\lambda_j}}{\sum_{k=1}^{J} e^{\lambda_k}}. \tag{3}$$

By scaling the relative importance of welfare attributes with the preference share method, the ratio of a given attribute preference share to another attribute preference share directly translates to their relative importance (Wolf and Tonsor, 2013). Utilizing the Krinsky–Robb (1986) method, we estimated confidence intervals for welfare attribute preference shares at the sample level. Relative significant differences between preferences for those attributes were assessed using overlapping confidence intervals (Olynk, Tonsor, and Wolf, 2010).

Preference shares were calculated at both the individual level and the sample (or population estimate) level to understand relative preferences and to allow us to assess correlations between preference shares and respondent demographics. A benefit of developing individual preference shares is the ability to assess consistency across preference elicitation methods for each respondent. From the industry perspective, an understanding of individual-level preferences is especially important when considering targeted marketing.

Comparing Methods of Ranking Preferences

Lagerkvist (2013) found that best–worst scaling methods were more effective at predicting actual consumer choices than a forced ranking of attributes and discussed the potential violation of transitivity between the two methods (i.e., discrepancies in the preference order). But even within the best–worst method, research suggests that the specific design of the best–worst questions can have an effect on the results (Byrd, Widmar, and Gramig, 2018). Given the importance of understanding consumer preferences for the laying-hen industry, consistency among the different preference elicitation methods to inform management, investment, and policy are key. Where discrepancies arise between methods, drawing conclusions about consumer preference or value placement is difficult. In such scenarios, the same individual may place value on different welfare attributes based on the question framing rather than an inherent value difference. By exploring multiple preference elicitation methods for welfare attributes in hen housing systems, this research bridges preference elicitation and welfare scoring to unite animal science expertise and consumer preference in a usable form.

We compared all four preference elicitation methods, resulting in four potential rankings: mean Likert, mean forced rank, mean of individual preference share estimates (using individual-specific coefficient estimates), and sample preference shares (using sample coefficient estimates). To convert the preference shares to a comparable ranking, values 1 through 7 were assigned to the largest to the smallest preference share. Where significant differences were not found, attributes were assigned the same rank-order value. Overlapping 95% confidence intervals were used to determine significant differences between attributes for the Likert, forced rank, mean individual preference, and sample preference shares.[3] For example, when an attribute that was deemed most important for any given rating method was not significantly different from the second most important attribute for that method, they both were ranked first. Where significant difference was detected, they were ranked first and second, respectively. In this way, the preference elicitation methods were standardized to a 1 through 7 ranking scale.

[3] Comparing 95% confidence intervals and examining overlap is more conservative than standard significance tests when the null hypothesis is true and falsely fails to reject the null hypothesis more frequently than the standard method when the null hypothesis is false (Schenker and Gentleman, 2001).

Attribute Preferences at the Individual Level

We utilized four criteria to compare the consistency of rankings at the individual level for the best–worst scaling and forced rank.[4] First, "absolute consistency," in which the order of forced rank and the ranking of mean preference shares were identical. Second, "within one," in which all forced rankings were identical or within one place of the ordered preference shares. For example, if an attribute was ranked second according to its individual preference share, it would satisfy the "within one" criteria if it was ranked first, second, or third in the ranking section. Theoretically, this criterion could still be met without a single ranking level matching between the individual preference share and forced rank. Third, "exact first and last," in which the first-ranked and seventh-ranked attributes were the same as the greatest and lowest preference share rankings. Based on Louviere (2004), Marley and Louviere (2005), and Finn and Louviere (1992), the human cognitive process is more adapted to picking extremes. As such, it would be expected to have a relatively high level of consistency in this criterion, looking at extremes only. And fourth, "top three and bottom three," in which the same sets of attributes in the top three rankings and bottom three rankings were the same but the order of those attributes did not matter. For example, this criterion could be met if a set of three attributes ranked first, second, and third by individual preference share were ranked third, second, and first, respectively, in the forced rank section. Using these comparison methods, we evaluated the degree of consistency of responses for each individual respondent. Potential limitations of this method are the inability to differentiate between individuals who have no preference or individuals who lack consistency in their preferences.

Consumer Based Welfare Scoring Index (CBWSI)

Animal welfare experts have developed scoring methods that are animal focused but do not consider consumer preferences for animal welfare and other production attributes (Coalition for Sustainable Egg Supply, 2015; American Veterinary Medical Association, 2008). This research connects best practices as determined by animal and poultry science experts with consumer preference in the areas of hen and worker welfare. By multiplying the preference shares for different attributes revealed by the best–worst design (Table 3) with the attribute scores assigned by CSES to the three hen housing types (Table 1), we created a unique score for welfare under each hen housing system. We denote this score the Consumer Based Welfare Scoring Index (CBWSI). The CBWSI takes the welfare scores assigned through the research conducted by animal welfare specialists and weights using individuals' revealed preferences.

Results

Summary Statistics

A total of 2,574 completed surveys were collected. Table 2 reports summary statistics for the survey respondents and compares them to U.S. Census Bureau population estimates. As the table reveals, the sample closely matched the U.S. population with respect to gender, age, income, education, and region of residence. Additionally, 86% of respondents were the primary shopper in the household making their views of egg housing practices particularly relevant for understanding preferences. The overwhelming majority (86%) of respondents consumed eggs at least every week, while only 3% did not consume eggs. Respondents reported generally low self-assessed knowledge of egg production, with 78% indicating neutral or lower. Finally, when asked what the most important reason for not

[4] At the individual-respondent level, the Likert scale does not allow for direct comparisons since individuals can assign the same score to any or all attributes. Thus, comparisons of Likert scale rankings to the other methods studied were reserved for the sample evaluations.

Table 2. Summary Statistics

	Survey Frequency (no. of respondents)	Survey (% respondents)	U.S. Census Bureau, 2016 (% population)
Gender			
Female	1,405	54.6	51.3
Male	1,169	45.0	48.7
Age			
18–24	190	7.4	12.6
25–34	522	20.3	17.8
35–44	474	18.4	16.4
45–54	487	18.9	17.4
55–64	449	17.4	16.5
65+	452	17.6	19.3
Income			
$0–$24,999	538	20.9	22.1
$25,000–$49,999	669	26.0	22.7
$50,000–$74,999	504	19.6	16.7
$75,000–$99,000	361	14.0	12.1
$100,000+	502	19.5	26.4
Geographic region			
Northeast	487	18.9	17.5
South	941	36.6	37.7
Midwest	574	22.3	21.1
West	572	22.2	23.7

making an egg purchase, the most common reason was price (42% of respondents), while only 7% of respondents selected factors associated with animal welfare.

Attribute Preferences

Table 3 presents the mean score, rank, or preference share results (including confidence intervals) for all four preference elicitation methods. The Likert scale mean scores were closely grouped around a score of 4.0, where a higher score indicates higher level of preference. For the forced rank, the averages were also clustered between 3.5 and 5, where a lower score indicates a more preferred rank. The sample best–worst preference shares reveal that hen mortality, stress, and behavior had percentage preference shares near 20% for the sample (e.g., 0.202 for hen behavior indicates a 20.2% preference share), while worker health and safety and cannibalism/aggression had about 15% shares. Feather condition was considered relatively less important, with only about a 5% preference share. For the mean individual preference share, the relative ranking was quite similar to the sample preference share. Note that the preference shares are cardinal, so the larger share indicates not just the sign but the relative magnitude of preference for that attribute.

Table 4 displays the statistically significant rank order of attributes by method. General patterns in the ranking of importance preferences can be discerned. One is that for the Likert, forced rank, and best–worst methods, hen behavior and hen stress were at least part of the top tier of attributes (i.e., most important). Hen mortality was tied for the most important attribute for the Likert and best–worst methods and the second most important using the forced rank method. The general pattern was that hen behavior, stress, and mortality were in the top tier of concerns. A close secondary tier of concerns were worker health and safety and hen cannibalism/aggression. Consistently, the

Table 3. Sample-Level Preference Elicitation Results

Attribute	Likert Score[a]	Forced Rank[b]	Sample Preference Share[c]	Mean Individual Preference Share[c]
Hen mortality	4.139	3.538	0.202	0.172
	(4.091–4.187)	(3.466–3.609)	(0.193–0.211)	(0.166–0.177)
Hen foot condition	4.046	4.721	0.064	0.050
	(3.998–4.095)	(4.653–4.788)	(0.062–0.067)	(0.048–0.051)
Hen behavior	4.128	3.453	0.189	0.214
	(4.080–4.176)	(3.377–3.529)	(0.179–0.199)	(0.206–0.222)
Hen stress	4.134	3.357	0.196	0.170
	(4.086–4.182)	(3.285–3.428)	(0.187–0.205)	(0.165–0.176)
Hen cannibalism/	4.117	4.077	0.146	0.124
aggression	(4.068–4.166)	(4.004–4.149)	(0.139–0.152)	(0.120–0.128)
Hen feather	3.94	4.973	0.046	0.034
condition	(3.890–3.991)	(4.905–5.041)	(0.044–0.048)	(0.034–0.035)
Worker health	4.100	3.883	0.157	0.236
and safety	(4.051–4.419)	(3.795–3.971)	(0.145–0.169)	(0.226–0.246)

Notes: Means are score for Likert, rank for forced rank, and preference share for best–worst. Numbers in parentheses are 95% confidence intervals.
[a] Likert scale was 1 = not important to 5 = extremely important.
[b] Forced rank required respondents to rank the attributes from 1 through 7.
[c] Preference shares derived from best–worst choices. Sample indicates the mean across all survey respondents, while individual were means across individual preference shares.

Table 4. Preference Elicitation Method Comparison[a]

Attribute	Likert	Forced Rank	Sample Preference Share	Individual Preference Share	Average Rank across Methods
Hen mortality	1	3	1	3	2.00
Hen foot condition	1	6	6	6	6.00
Hen behavior	1	1	1	2	1.25
Hen stress	1	1	1	3	1.50
Hen cannibalism/aggression	1	5	4	5	3.75
Hen feather condition	7	7	7	7	7.00
Worker health and safety	1	4	4	1	2.50

Notes: [a] Values in the table are statistically significant rank order of attributes for each method at mean.

least important attributes were hen foot condition (consistently sixth) and hen feather condition (consistently last).

Likert scaling provided little differentiation among attributes. Many respondents, 894 in total, answered "I don't know" rather than scoring at least one of the welfare attributes in the Likert scale section of the survey.[5] Even after dropping these responses, the only attributes that were ranked lower were foot and feather condition. The result is consistent with results from Cohen

[5] Even after eliminating these uncertain respondents, the Likert scale did not statistically differentiate among the top five welfare attributes. It would be expected that statistically significant differences between attributes would become more prevalent after eliminating uncertain respondents.

(2003), Cohen and Orme (2004), and Jaeger et al. (2008) that Likert scales are a less discriminatory preference elicitation method.

The largest discrepancy in rankings across methods was for the worker health and safety attribute, which moved from the most important attribute under the mean of individual shares to the fourth-ranked attribute for the sample preference shares. Worker health and safety preference share moves from 15.7% under the sample level model to 23.6% under the mean of individual shares. Further, the confidence intervals generated for the best–worst from the Krinsky–Robb method conclude that worker health and safety was statistically less important than mortality at the 95% level, while overlapping confidence intervals of these attributes using the mean of individual shares suggest the opposite. This discrepancy is especially important because the variation in CSES worker health and safety scores between housing systems cause this attribute to play a key role in determining the most preferred housing system.

Individual-Level Consistency

Comparing individual preference shares and individual forced ranks using each of the consistency criteria discussed in the methods section revealed very low consistency levels. Some inconsistency was expected given the generally low self-assessed knowledge of farming practices used in egg production indicated by survey respondents (77% neutral or lower). However, the absolute consistency level of under 2% of respondents was still surprising. Even the more generic criteria indicate relatively low levels of consistency, with only 20% of respondents demonstrating consistency in their absolute top and bottom priorities (first and last method) and just under 10% consistency according to the within one (9.98%) and top three and bottom three (9.95%) methods. Because statistical significance at the individual level cannot be established (as it is a single observation), consistency across methods provides a means of gauging relatively more or relatively less definitive preference for a given attribute or set of attributes. These results suggest uncertainty about priorities for welfare attribute preference. However, the statistically significant groupings of attributes at the sample level are mirrored at the individual level, with increasing consistency in groupings and at the extremes. For industry, consistency in consumer preference would suggest priority marketing categories or labeling considerations. The low levels of consistency in this research would make such segmentation challenging. Further, consistency levels had little correlation with demographic categories such as education level, gender, or knowledge about animal welfare attributes.[6] From the perspective of legislation, successful long-term policy depends on known and consistent preferences. These generally inconsistent results suggest that U.S. consumers may not have well-defined, consistent, stable preferences for welfare attributes. This inconsistency is potentially problematic when mandated long-term investment in hen housing leads to inverse welfare effects on hens versus workers. If uncertain preference leads to a new ballot initiative or legislation, that seemingly subtle change in relative preference for a given welfare category can actually change the housing system that is required to enact those priorities.

Consumer Based Welfare Scoring Index (CBWSI)

Combining the individual preference shares (Table 3) with the scoring rubric of hen housing systems developed by CSES (Table 1), Table 5 reports ratings for housing systems using a combination of documented welfare effects and consumer preferences. Utilizing the sample preference shares, the enriched colony housing system is superior for the average U.S. consumer.[7] The same is true for the even-weighting scheme, although the magnitude of individual attributes shifts by the difference

---

[6] Self-assessed knowledge on animal welfare was indicated by “I don’t know” answers in the Likert section of the survey.

[7] Because conventional housing was the baseline for all CSES scoring, it yields a 0 value in this scenario regardless of the preference share weightings and is not included in the table. However, the conventional cage relative score can still be compared to the other housing systems by looking at their departures from 0.

Table 5. Consumer-Based Welfare Scoring Index (CBWSI)[a]

	Weighted by Sample Preference Share		Weighted Evenly	
	Enriched Colony	**Cage-Free Aviary**	**Enriched Colony**	**Cage-Free Aviary**
Mortality	0.00	−0.40	0.00	−0.29
Foot condition	0.20	0.00	0.14	0.00
Behavior	0.40	0.61	0.29	0.43
Stress	0.00	0.00	0.00	0.00
Cannibalism/aggression	−0.20	−0.61	−0.14	−0.43
Feather condition	0.00	0.20	0.00	0.14
Worker health and safety	0.00	−0.23	0.00	−0.16
CBWSI[b]	0.40	−0.43	0.29	−0.31

Notes: [a] Values in the table are mean preference share values (from Table 3) multiplied by CSES score relative to conventional cages (from Table 1).
[b] The CBWSI Score for each housing system was calculated by multiplying each of the CSES scores by the respective preference share for the given attribute and then summing across all attributes.

between their preference share value and 0.14 (one-seventh) times that attributes welfare score. Cage-free aviaries have a negative rating under each weighting scheme, indicating that they are inferior to enriched colony housing and conventional cage housing for the average consumer when considered as the weighted sum of attribute preferences. However, there were survey respondents whose individual CBWSI suggested that cage-free aviaries were superior given their individual preference shares for the seven attributes. For that to happen, those individuals had very high preference shares for behavior that overwhelmed the small importance they placed on the other negatively rated attributes. Conversely, respondents who had high preference shares for worker health and safety had much larger CBWSI values for conventional cage and enriched colony housing due to the high negative value for that attribute under cage-free aviaries.

Discussion

Preference Elicitation Methods

Although sample-level preferences for attributes were similar across elicitation methods, individual attribute ranking were quite inconsistent. Given the ease and increasing popularity of online survey data collection, it is important to consider the variability introduced simply by the preference elicitation method chosen. Whether it is the egg industry using stated preferences for investing in new hen housing design or government developing policy, the preference elicitation method may have a dramatic effect on the course of action that is chosen. This is not a new issue, as Lagerkvist (2013) found similar issues when comparing best–worst scaling and forced rank. Even though best–worst provided statistically superior differentiation of attributes (Lagerkvist, 2013), it could be that the validity of that differentiation exists only for respondents who are well-informed about the topic and the specific attributes in question. By forcing a trade-off for respondents who have no preference or who are not well-informed enough to form a preference about a given set of attributes, some aspect of the differentiation may be artificially imposed. Although the lack of a forced trade-off is a weakness of the Likert scaling method, it does allow for even scaling or neutral selections for all attributes in the face of uncertainty. Cohen (2003) found that the best–worst preference elicitation method took the average respondent twice as long to complete as the rating task and went on to suggest that the strength in the method may simply be a product of the increased effort put forth. As an extension for future research, it may be important to repeat the less arduous preference tasks such as the Likert scale. This would filter respondents who are inconsistent even when answering the same questions twice and induce greater cognitive effort for the simpler method. Regardless of the approach, further research should compare and contrast preference elicitation methods to better understand the observed inconsistencies in this and other research.

Revealed Preference versus Industry Direction

The CBWSI results indicate that cage-free hen housing systems are inferior to both enriched colony and conventional production systems according to consumer preferences for welfare attributes. Cage-free egg production has grown tremendously in recent years and is expected to continue to do so given commitments to go cage-free by large retailers such as Walmart and McDonalds (Walmart, 2016; Jargon and Beilfuss, 2015).

The interest in cage-free aviary–produced eggs has grown so substantially that the USDA's Agricultural Marketing Service (AMS) began releasing a new market report dedicated to cage-free eggs in September 2016. According to the March 2018 version of this report, there were 51,857,000 cage-free laying hens in the United States (U.S. Department of Agriculture, 2018b), which represents 16% of the total U.S. flock of 324 million laying hens (U.S. Department of Agriculture, 2018b). Although the percentage of the total flock is still relatively small, cage-free production grew 120% since January 2016 and 38% since January 2017 (U.S. Department of Agriculture, 2018b; American Egg Board, 2018). Given the commitments by large retailers previously discussed, this growth is likely to continue. Despite the buzz around cage-free eggs, there are not uniform or official definitions for the term. The USDA AMS defines cage-free for its labeling requirements as follows:

> Eggs packed in USDA grade marked consumer packages labeled as cage free must be produced by hens housed in a building, room, or enclosed area that allows for unlimited access to food, water, and provides the freedom to roam within the area during the laying cycle (U.S. Department of Agriculture, 2016).

A more extensive definition was approved by the United Egg producers in May 2017:

> Cage-free eggs are laid by hens that are able to roam vertically and horizontally in indoor houses, and have access to fresh food and water. Cage-free systems vary from farm to farm, and can include multi-tier aviaries. They must allow hens to exhibit natural behaviors and include enrichments such as scratch areas, perches and nests. Hens must have access to litter, protection from predators and be able to move in a barn in a manner that promotes bird welfare (United Egg Producers, 2017).

Certified Humane, an independent certifying agency, goes a step further and outlines cage-free certification requirements in a full list of categories from flooring and enhancements to air and light (Humane Farm Animal Care, 2017). With the growing interest in cage-free eggs and the lack of standardized definitions, it may be more productive to think of consumer preferences for housing systems as a bundled set of attributes rather than in terms of variable definitions. Given the results of this research, public preferences for bundles of welfare attributes are more in line with enriched colony hen housing systems than with the recently popularized move toward cage-free production. This disconnect might result from a lack of information.

One potential explanation for this result is a lack of consumer knowledge about the welfare effects of various hen housing systems. Not only was cage-free aviary rated as the worst system across hen welfare scores, but it also requires 179% higher capital costs per dozen eggs than conventional cage production, compared to 106% higher for the enriched colony per dozen eggs (Coalition for Sustainable Egg Supply, 2015). To align consumer expectations for housing systems with the animal welfare reality of those systems, there must be greater understanding and transparency of egg housing-system labels. Rather than a simple label of housing type, it may be more beneficial to move toward a CBWSI that can be easily calculated from a series of questions along with animal welfare scores on a given product. The average U.S. consumer believes that purchasing cage-free aviary eggs instead of conventional cage eggs results in an improvement in all aspects of hen and worker welfare, while animal welfare experts would suggest enriched colony systems are more in line with this expectation. In future research using visual aids, it may be important to determine whether this result is an anomaly or if improved consumer understanding of the system as whole yields preferences consistent with this aggregated index of welfare attributes.

Conclusions

The most important laying-hen welfare attributes for the average U.S. consumer were hen behavior, hen stress, and hen mortality. Cannibalism/aggression and worker health and safety were of intermediary importance, and least important were hen feather condition and hen foot condition. We combined welfare scores from animal science experts and revealed preferences from consumers into a Consumer Based Welfare Scoring Index (CBWSI), which suggests that enriched colony hen housing is the superior system. Using this same metric, cage-free housing would be rated inferior to conventional cages. The revealed preferences for housing systems via the CBWSI are in direct contrast to the rapid shift toward cage-free housing. The contrast between revealed preferences for hen housing and the growing demand for cage-free eggs demonstrates the need for greater education of the average U.S. consumer regarding hen housing systems and greater transparency in animal welfare labeling.

[First submitted May 2018; accepted for publication August 2018.]

References

American Egg Board. "Industry Overview." 2018. Available online at https://www.aeb.org/farmers-and-marketers/industry-overview.

American Veterinary Medical Association. "AVMA Issues - A Comparison of Cage and Non-Cage Systems for Housing Laying Hens." 2008. Available online at https://www.avma.org/KB/Resources/ Reference/AnimalWelfare/Pages/AVMA-issues-A-Comparison-of-Cage-and-Non-Cage-Systems-for-Housing-Laying-Hens.aspx.

Bir, C., N. Widmar, and C. Croney. "Stated Preferences for Dog Characteristics and Sources of Acquisition." *Animals* 7(2017):59. doi: 10.3390/ani7080059.

Blatchford, R. A., R. M. Fulton, and J. A. Mench. "The Utilization of the Welfare Quality® Assessment for Determining Laying Hen Condition across Three Housing Systems." *Poultry Science* 95(2016):154–163. doi: 10.3382/ps/pev227.

Bruilliard, K. "Many People Think a Cage-Free Life Is Better for Hens. It's Not that Simple." *Washington Post* (2017). Available online at https://www.washingtonpost.com/news/animalia/wp/2017/06/15/many-people-think-cage-free-life-is-better-for-hens-its-not-that-simple/.

Byrd, E. S., N. J. O. Widmar, and B. M. Gramig. "Presentation Matters: Number of Attributes Presented Impacts Estimated Preferences." *Agribusiness* 34(2018):377–389. doi: 10.1002/agr.21527.

California Legislature. "Prevention of Farm Animal Cruelty Act. CA hsc. §§25990–25994." 2008. Available online at https://leginfo.legislature.ca.gov/faces/codes_displayText.xhtml?lawCode=HSC&division=20.& title=&part=&chapter=13.8.&article.

Carson, R. T., and J. J. Louviere. "A Common Nomenclature for Stated Preference Elicitation Approaches." *Environmental and Resource Economics* 49(2011):539–559. doi: 10.1007/s10640-010-9450-x.

Coalition for Sustainable Egg Supply. *Final Research Results Report*. Gladstone, MO: Center for Food Integrity, 2015. Available online at http://www2.sustainableeggcoalition.org/final-results.

Cohen, S., and B. Orme. "What's Your Preference?" *Marketing Research* 16(2004):32–37.

Cohen, S. H. "Maximum Difference Scaling: Improved Measures of Importance and Preference for Segmentation." Research Paper Series, Sawtooth Software, Sequim, WA, 2003. Available online at https://www.sawtoothsoftware.com/download/techpap/maxdiff.pdf.

Finn, A., and J. J. Louviere. "Determining the Appropriate Response to Evidence of Public Concern: The Case of Food Safety." *Journal of Public Policy & Marketing* 11(1992):12–25. doi: 10.1177/074391569201100202.

Flynn, T. N., J. J. Louviere, T. J. Peters, and J. Coast. “Best–Worst Scaling: What It Can Do for Health Care Research and How to Do It.” *Journal of Health Economics* 26(2007):171–189. doi: 10.1016/j.jhealeco.2006.04.002.

Hein, K. A., S. R. Jaeger, B. Tom Carr, and C. M. Delahunty. “Comparison of Five Common Acceptance and Preference Methods.” *Food Quality and Preference* 19(2008):651–661. doi: 10.1016/j.foodqual.2008.06.001.

Humane Farm Animal Care. “What Makes Certified Humane Cage-Free Eggs Different from Other Cage-Free Eggs?” 2017. Available online at https://certifiedhumane.org/makes-certified-humane-cage-free-eggs-different-cage-free-eggs/.

Jaeger, S. R., A. S. JÃÿrgensen, M. D. Aaslyng, and W. L. Bredie. “Best–Worst Scaling: An Introduction and Initial Comparison with Monadic Rating for Preference Elicitation with Food Products.” *Food Quality and Preference* 19(2008):579–588. doi: 10.1016/j.foodqual.2008.03.002.

Jargon, J., and L. Beilfuss. “McDonald’s Continues Image Shift with Move to Cage-Free Eggs in North America.” *Wall Street Journal* (2015).

Krinsky, I., and A. L. Robb. “On Approximating the Statistical Properties of Elasticities.” *Review of Economics and Statistics* 68(1986):715–719. doi: 10.2307/1924536.

Lagerkvist, C. J. “Consumer Preferences for Food Labelling Attributes: Comparing Direct Ranking and Best–Worst Scaling for Measurement of Attribute Importance, Preference Intensity and Attribute Dominance.” *Food Quality and Preference* 29(2013):77–88. doi: 10.1016/j.foodqual.2013.02.005.

Lay, D. C., R. M. Fulton, P. Y. Hester, D. M. Karcher, J. B. Kjaer, J. A. Mench, B. A. Mullens, R. C. Newberry, C. J. Nicol, N. P. O’Sullivan, and R. E. Porter. “Hen Welfare in Different Housing Systems.” *Poultry Science* 90(2011):278–294. doi: 10.3382/ps.2010-00962.

Likert, R. “A Technique for the Measurement of Attitudes.” *Archives of Psychology* 22(1932):55.

Louviere, J. J. “Best–Worst Scaling. Theory and Example Applications.” 2004. Invited seminar, University of Auckland, New Zealand, Department of Marketing, March 26.

Lusk, J. L., and B. C. Briggeman. “Food Values.” *American Journal of Agricultural Economics* 91(2009):184–196. doi: 10.1111/j.1467-8276.2008.01175.x.

Marley, A. A. J., and J. J. Louviere. “Some Probabilistic Models of Best, Worst, and Best–Worst Choices.” *Journal of Mathematical Psychology* 49(2005):464–480.

Matthews, W. A., and D. A. Sumner. “Effects of Housing System on the Costs of Commercial Egg Production.” *Poultry Science* 94(2015):552–557. doi: 10.3382/ps/peu011.

Mayer, A. “Housing Hens Ruffles Some Feathers in the Egg Industry.” *Harvest Public Media, NPR* (2015).

Michigan Legislature. “Michigan Public Act 117. MCL 287.746.” 2009. Available onilne at http://legislature.mi.gov/doc.aspx?mcl-287-746.

Netherlands Welfare Quality® Consortium. “Welfare Quality®, Assessment Protocol for Poultry (Broilers, Laying Hens).” 2009. Lelystad, Netherlands.

Olynk, N. J., G. T. Tonsor, and C. A. Wolf. “Consumer Willingness to Pay for Livestock Credence Attribute Claim Verification.” *Journal of Agricultural and Resource Economics* 35(2010):261–280.

Schenker, N., and J. F. Gentleman. “On Judging the Significance of Differences by Examining the Overlap Between Confidence Intervals.” *American Statistician* 55(2001):182–186. doi: 10.1198/000313001317097960.

State of Florida. “Florida Constitution, Article 10 §21: Limiting Cruel and Inhumane Confinement of Pigs during Pregnancy.” 2002. Available online at http://www.leg.state.fl.us/statutes/index.cfm?submenu=3#A10S21.

Swanson, J. C., Y. Lee, P. B. Thompson, R. Bawden, and J. A. Mench. “Integration: Valuing Stakeholder Input in Setting Priorities for Socially Sustainable Egg Production.” *Poultry Science* 90(2011):2110–2121. doi: 10.3382/ps.2011-01340.

United Egg Producers. “Statement to Stakeholders: UEP Definition of Cage-Free Egg Production.” 2017. Available online at https://unitedegg.com/statement-to-stakeholders-uep-definition-of-cage-free-egg-production/.

U.S. Department of Agriculture. *USDA Grademarked Product Label Checklist. Agricultural Marketing Service*. Washington, DC: U.S. Department of Agriculture, Agricultural Marketing Service, 2016. Available online at https://www.ams.usda.gov/sites/default/files/media/USDA Grademarked Product Label Submission Checklist.pdf.

———. *February Chickens and Eggs Report*. Washington, DC: U.S. Department of Agriculture, National Agricultural Statistics Service, 2018a. Available online at https://web.archive.org/web/20180329184855/usda.mannlib.cornell.edu/usda/current/ChicEggs/ChicEggs-03-23-2018.pdf.

———. *Monthly USDA Cage-Free Shell Egg Report*. Washington, DC: U.S. Department of Agriculture, Agricultural Marketing Service, 2018b. Available online at https://www.ams.usda.gov/mnreports/pymcagefree.pdf.

Walmart. “Walmart U.S. Announces Transition to Cage-Free Egg Supply Chain by 2025.” 2016. Available online at https://news.walmart.com/_news_/news-archive/2016/04/05/walmart-us-announces-transition-to-cage-free-egg-supply-chain-by-2025.

Watnick, V. “The Business and Ethics of Laying Hens: California’s Groundbreaking Law Goes into Effect on Animal Confinement.” *Boston College Environmental Affairs Law Review* 43(2016):45.

Whoriskey, P. “More than a Million Hens, Filling Barns at Three per Square Foot. And Yes, They’re USDA Organic.” *Chicago Tribune* (2017).

Willits, F. K., G. L. Theodori, and A. E. Luloff. “Another Look at Likert Scales.” *Journal of Rural Social Sciences* 31(2016):126–139.

Wolf, C. A., and G. T. Tonsor. “Dairy Farmer Policy Preferences.” *Journal of Agricultural and Resource Economics* 38(2013):1–15.

Yilmaz Dikmen, B., A. İpek, U. Şahan, M. Petek, and A. Sözcü. “Egg Production and Welfare of Laying Hens Kept in Different Housing Systems (Conventional, Enriched Cage, and Free Range).” *Poultry Science* 95(2016):1564–1572. doi: 10.3382/ps/pew082.

Appendix A: Coefficient Estimates

Variable	Multinomial Logit	Random Parameters Logit	
	Coeff.	Coeff.	Std. Dev.
Mortality	1.0147***	1.4817***	1.1051***
	(0.0198)	(0.0313)	(0.0317)
Foot condition	0.2592***	0.3335***	0.2842***
	(0.0192)	(0.0228)	(0.0474)
Behavior	0.9982***	1.4146***	1.5671***
	(0.0198)	(0.0361)	(0.0365)
Stress	1.0073***	1.4495***	1.1143***
	(0.0198)	(0.0309)	(0.0327)
Cannibalism/aggression	0.8249***	1.1536***	0.9867***
	(0.0196)	(0.0286)	(0.033)
Worker health and safety	0.8249***	1.2292***	2.3230***
	(0.0196)	(0.0449)	(0.0499)
Feather condition	Baseline	Baseline	Baseline

Notes: Triple asterisks (***) indicate significance at the 1% level. Standard errors in parentheses.