Consistency and cognitive influences on the expert judgement of Environmental Health Officers

An Experimental Study

Dr. Jeroen Nieboer (Department of Social Policy)
Dr. Tom Reader (Department of Social Psychology)
with Ms. Delphine Leardini

FSA GSR Profession
Food Standards Agency
FS516018
December 2015

TABLE OF CONTENTS

EXECUTIVE SUMMARY.

This report detailed the results of an experimental study on the consistency and variance of the assessments of Environmental Health Officers (EHOs), conducted at the London School of Economics and Political Science in March 2015. The study was commissioned by the Food Standards Agency (call ref. #FS516018).

Introduction

- This research project investigated psychological factors that lead EHOs to make inconsistent judgements about catering outlets. It draws on research from the fields of 'cognitive biases' and 'naturalistic decision making' to understand decision making on food hygiene
- These are fields of study within psychology, and they examine how variations in decision-making arise from heuristics (rules and habits that influence decision-making) and cognitive bias (errors that arise from a heuristic being applied incorrectly)
- Field experiments are generally considered to be the 'gold standard' method of investigating cognitive bias, and to investigate variance in decision-making by EHOs and the extent to which psychological factors explain this variance, decision-making on food hygiene was studied through a series of experiments

Method

- Based within a food and beverage outlet at the London School of Economics and Political Science (LSE) campus, two 'simulated kitchens' were created, and EHOs were asked to inspect and assess the two kitchen environments
- Scenarios were designed to examine the impact of three specific biases upon the decision making of EHOs: anchoring, confirmation bias, and overconfidence
- The simulated kitchens replicated aspects of a real catering environment (e.g. equipment, facilities food preparation, catering staff, food storage), and contained a number of positive (e.g. food storage) and negative informational cues (e.g. out-of-date records) that were expected to influence ratings in food hygiene (environmental health) inspections

- To assess whether an ‘anchoring bias’ influences assessments of food safety, we examined whether the assessments of the experimental catering facilities by inspectors are in-part determined by information on a previous inspection
- To assess if ‘confirmation bias’ influences assessments of food hygiene, we examined whether the decision-making of EHOs is influenced by an early informational cue presented during an inspection
- To assess ‘over-confidence bias ’, we examined whether high or low confidence in EHOs (as measured by a post-study survey on self-assessed confidence and confidence on a number of knowledge items) results in differential assessments of catering facilities

Results and Analysis

- The experimental results suggest that inspectors are not susceptible to anchoring bias on previous ratings, nor to confirmation bias triggered by the order in which information was presented to them
- Yet, considerable variance in ratings was found, and this was primarily explained by differences in scoring confidence in management.
- Through post-study surveys, EHOs reported basing inspection decisions on a range of cues (e.g. chef knowledge, documentation), and this was particular to each EHO.
- This, along with limitations in the study design, likely explains the lack of experimental effect in this study

Recommendations

- Future research and training may wish to focus more upon the patterns and cues that EHOs use to make food safety decisions, rather than the generic biases that may influence these.
- Future research may wish to leave the laboratory, and focus on heuristics and biases in the real world.
- Despite recognising some of the limitation of laboratory-style environments, we think they may be appropriate for training inexperienced EHOs.

MAIN REPORT.

1. Introduction

This research project investigated psychological factors that lead Environmental Health Officers (henceforth EHOs) to make inconsistent judgements about food and beverage outlets. It draws on research from the fields of 'cognitive biases' and 'naturalistic decision making' to understand decision making on food hygiene (Tversky & Kahneman, 1981; Zsambok & Klein, 2014). These scholarly fields examine how aspects of situational constraints, information processing, and individual expertise influence decision making in laboratory and real-life situations. Whilst analogous research has investigated inconsistent decision making in the fields of medical and criminal justice, little decision making research has been conducted in the domain of food safety and hygiene enforcement.

The EHOs that participated in our study were employed by Local Authorities throughout the United Kingdom. Their key responsibility is to ensure the safety, health and hygiene of food and beverage facilities in their local area. These facilities range from public institutions, such as schools and hospitals, to private companies, such as restaurants, catering companies and production environments. The scenario used in our experimental study, a kitchen preparing food for sale on a university campus, is thus representative of a type of outlet that an EHO may encounter in their everyday work. Our study focuses on the decision-making of EHOs during and immediately after an inspection of the outlet. The inspection was modelled on the experience of shadowing an EHO on duty, which a member of our research team undertook prior to the design of the study.

Psychology research examining inconsistencies in decision-making tends to not assume that individuals are 'poor' decision-makers, but that variations in decision-making arise from heuristics (rules and habits that influence decision-making) and cognitive bias (errors that arise from a heuristic being applied incorrectly). Broadly, the academic field of cognitive bias investigates inconsistencies in decision-making that occur due to the subconscious influence of social and emotional factors, and the way information is processed (e.g. ignored or prioritised) and evaluated (e.g. assessing probability). Whilst heuristics allows decision-makers to be highly efficient (e.g. for processing complex information and making decisions quickly), the biases that underlie them can also create the possibility for error. For example,

three of the most commonly examined cognitive biases are confirmation bias, anchoring bias, and confidence bias (Kahneman & Tversky, 1996). Confirmation bias relates to decision-makers collecting information to confirm a previously made decision, and ignoring information that indicate another decision to be more appropriate. Anchoring bias refers to the tendency of a decision maker to heavily base decision-making on an initial piece of information presented to them. Confidence bias relates to decision makers believing they are more accurate in decision-making than they are objectively.

Researchers have extensively investigated the influence of cognitive bias upon decision-making in a range of professional settings: for example in medicine (e.g. clinician errors), finance (e.g. investment decisions), aviation (e.g. pilot judgements), and forensic science (e.g. assessing criminal evidence). To do this, researchers have typically created experimental 'field' studies, whereby they attempt to simulate a real-world environment in the laboratory (e.g. a simulated cockpit, medical environment, technical setting) (Kersholt et a., 2010; Tschan et al., 2009). This is generally considered to be the 'gold standard' way of examining bias amongst expert professionals because:

i. decision-makers are placed in naturalistic situation that reflects a working environment in which decisions are made,
ii. researchers have a high degree of control for creating the environment (e.g. in placing and designing information cues),
iii. decision scenarios can be standardised so that they are consistent across all study participants.

This means that variations observed in decision-making can be explained by an experimental manipulation intended to examine a cognitive bias, and not variation in the study setting. Such work has led to insights on error and judgement amongst expert professionals (e.g. the influence of confirmation bias in patient diagnosis errors), and has contributed to the design of interventions for improving decision-making (Crichton et al., 2000; Croskerry, Singhal & Mamede, 2013).

Thus, an insightful and practical way to examine the consistency of decision-making amongst EHOs is to investigate variations in decision-making within a simulated experimental environment. Furthermore, to understand factors that might generate variations in decision-

making EHOs, the cognitive bias approach appears particularly useful. In the current study we examine EHO decision-making in a field laboratory, and investigate the influence of the three most commonly investigated and detected cognitive biases (confirmation bias, anchoring bias, confidence bias) upon EHO assessments of simulated kitchen environments. For example, confirmation bias has been much studied in relation to case studies of poor decision-making in policy and government (Janis, 1972; Cannon-Bowers & Salas, 1998), anchoring biases in medical misdiagnosis (Croskerry, 2003), and over-confidence in financial decision-making (Bazerman & Moore, 2012). They are cited as especially important for understanding erroneous decision-making in field environments, and thus we examine them in the current study (Kahneman, 2011). Furthermore, since these biases are related to the processing of information, we judged them particularly important in the context of EHO judgements. Because EHOs have to process and interpret a lot of information in a relatively short space of time, cognitive biases are more likely to arise and affect judgements.

To investigate variance in decision-making by EHOs and the extent to which psychological factors explain this variance, decision-making on food hygiene was investigated through a series of experiments. A cohort of EHOs from across the United Kingdom were invited to participate in two field-based experiments. Based within a food and beverage outlet at the London School of Economics and Political Science (LSE) campus, two 'simulated kitchens' were created, and participating EHOs were asked to inspect these outlets. The simulated kitchens replicated aspects of a real catering environment (e.g. equipment, facilities food preparation, catering staff, food storage), and contained a number of positive (e.g. food storage) and negative informational cues (e.g. out-of-date records) that were expected to influence ratings in food hygiene (environmental health) inspections. Furthermore, scenarios were designed to examine the impact of three specific biases upon the decision making of EHOs: anchoring, confirmation bias, and overconfidence. Through examining the decision-making of EHOs within the simulated kitchens, the research aimed to develop insight on the psychological factors that influence assessments on the safety of food being delivered by catering establishments.

The rest of the report is organised as follows. Section 2 outlines the study research questions; section 3 describes the experimental methodology used to test these research questions; section 4 reports the experimental study results; section 5 contains supplementary analyses of the qualitative data and section 6 concludes.

2. Research questions

This research study had EHOs participating in two experimental catering scenarios in order to investigate the following research questions (RQs).

2.1. Experimental questions

RQ1. To what extent are EHOs consistent in the assessments they make of FBOs?
Studies of decision-making in most professional environments (e.g. finance, medicine, military, aviation) show considerable variance in how professionals assess and judge a problem or scenario (Cannon-Bowers & Salas, 1998; Bazerman & Moore, 2012). The starting premise for this investigation was that EHOs might, consistent with research on decision-making in other domains of expert knowledge, also show variation in their assessments of the safety and hygiene of food and beverage outlets. Put simply, the food hygiene assessment given to a catering outlet can sometimes depend on the officer who is conducting the assessment, and how they perceive, respond, and interpret information about food hygiene within a catering outlet. Whilst relying on experience and discretion in decision-making is critical to ensuring the effectiveness and skills of EHOs, too much variance in the outcomes of safety inspections is not optimal as it could indicate that standards for safety are not being applied appropriately or fairly. This observation corresponds with the literature on expert decision making, which shows that experts become highly reliant on mental shortcuts, expertise, and heuristics when decision making (Crosskerry & Norman, 2008; Klein, 1993). This can result in experts showing inconsistent patterns of decision making when faced with similar problem scenarios, and research in fields such as medicine, aviation, finance, and law has attempted to identify where, when, and why variance in expert decision-making occurs (and to develop interventions to ensure optimal decision-making). Thus, we predict EHOs will show considerable variance in their inspection of catering facilities, and examine this through the results of their inspections for the two LSE experimental kitchens.

RQ2. To what extent are EHOs influenced by cognitive biases in the assessments they make of catering outlets?
To understand the variance in the assessments of catering outlets made by EHOs (and provided there is variance), we invoked seminal scholarly works from the cognitive biases

literature (Simon, 1990; Tversky & Kahneman., 1981), and in short they indicate that human beings apply a number of sub-conscious strategies when making decisions (e.g. preferences for information, applying probability calculations). These strategies allow us to be highly effective in decision making (e.g. quick, able to handle multiple decisions), but are also subject to error. In particular, environmental factors (e.g. how information is presented) and individual factors (e.g. personal knowledge and expertise, habits) can 'invoke' a bias, and these biases have been shown to be particularly significant in organisational decision-making (Bazerman & Moore., 2012). In this study, we focus on three of the most common cognitive biases: anchoring, confirmation bias, and over-confidence.

RQ2a. Does 'anchoring' influence food hygiene judgements of EHOs?

Anchoring refers to the tendency of a decision maker to resolve ambiguity or uncertainty by taking an earlier judgment as a starting point. Subsequent information is then viewed through the lens of that anchor (Tversky & Kahneman, 1974). To examine whether anchoring influences assessments of food safety, we examine whether the assessments of the experimental catering facilities by EHOs are in-part determined by information on a previous inspection (i.e. a good or poor previous assessment, which serves as an anchor). We hypothesise that where the anchor is 'negative' (i.e. the catering facility was awarded a low score previously), the assessments of the catering facility will also be low (in comparison to the 'positive' anchor group).

RQ2b. Does 'confirmation bias' influence food hygiene judgements of EHOs?

Confirmation bias refers to the tendency to seek or interpret information that confirms the decision maker's preconceptions (Nickerson, 1998; Oswald, Grosjean, 2004). It is similar to anchoring, yet whilst anchoring refers to how an earlier judgement shapes decision making, confirmation bias refers to an in-situ and quick decision that is then justified through the gathering of information that confirms it. To examine if confirmation bias influences assessments of food hygiene, we examine whether the decision-making of EHOs is influenced by an early informational cue presented during an inspection (i.e. indicating the catering facility to have poor or good hygiene standards), which then shapes the strategy used to assess the catering facility (and the final assessment). We hypothesise that where the confirmation bias is 'negative' (i.e. the first informational cue presented indicates poor safety standards), the assessments of the catering facility will also be low (in comparison to the 'positive' confirmation bias group).

RQ2c. Does ‘over-confidence’ influence food hygiene inspections?

Overconfidence refers to the situation in which a person’s confidence in his or her judgements than justified (Thaler, Sunstein, 2008), a phenomenon commonly found in experts (Tetlock, 2005). In terms of food hygiene, over-confidence of EHOs (e.g. in the accuracy of their judgement, or how closely their judgements match those of their colleagues) might be expected to be associated with assessments of catering facilities that are systematically more extreme or variable. To examine over-confidence, we examine whether high or low confidence in EHOs (as measured by a post-study survey on self-assessed confidence and confidence on a number of knowledge items) results in differential assessments of catering facilities.

2.2. Demographic factors

Alongside the key research questions outlined above, a number of other issues emerge. In particular, these relate to the role of individual differences that might influence how EHOs assess catering facilities, for example experience or personality. We have not outlined specific research questions for these, because the might be expected to influence decision making across all experimental scenarios. To this effect, information about the individual differences of EHOs will be included as ‘control variables’ in the analysis of the research questions. To summarise, the individual differences that might be expected to influence food hygiene inspections include (and were measured through a post-study questionnaire):

Years of experience as an EHO. Naturalistic decision making research indicates that ‘expert knowledge’ on decision making is developed through experience (and being exposed to a variety of situations and scenarios), but that counter-intuitively this experience can make decision makers more reliant on a limited set of cues for making decisions (meaning they are more susceptible to cognitive bias than novices, who are methodical in decision making) (Crosskerry, 2003; Kahneman, 2011). It thus might be expected that experienced participants will be more likely to be influenced by the experimental manipulations outlined above.

Knowledge on awarding safety ratings. Research on cognitive biases indicates that ‘representative knowledge’ is a key influencer of decision making (Gigerenzer & Gaissmaier, 2011; Kahneman & Frederic, 2002). In short, our knowledge of past distributions and

probability shapes future decision making (e.g. of the number of catering facilities that receive a 5* assessment will distribute how many 5* awards an EHO thinks is 'normal' to award). For the current study this is relevant because what the EHOs consider to be 'normal' in terms of awarding food hygiene ratings (e.g. that over 50% of facilities are rated 5*) might shape their assessments in the experiment. For example, participants who believe that low food hygiene ratings are commonly awarded may be more likely to be conservative in their assessments of catering facilities. We thus examine the influence of knowledge about the awarding of safety assessments upon ratings made in the experimental scenarios.

Conscientiousness. Decision-making research indicates that individual differences can be quite influential in terms of shaping behaviour (Johnson & Weber, 2009; Lepine & Van Dyne, 2001). In particular, for tasks that require the careful evaluation of information, high conscientiousness people are more likely to systematically evaluate a decision scenario and reach a judgement based on this. Low conscientiousness people are more likely to reach a quick decision, and to be less thorough in evaluation (although the empirical evidence on this is mixed: LePine, Colquitt, & Erez, 2000). Thus, it might be expected that participants who are less conscientious will be more likely to be influenced by the experimental manipulations outlined above.

Age and gender. Within experimental research, age and gender are commonly measured as a control variable. For this study we make no specific hypotheses relating to these variables, and will include them as control variables.

3. Methods

The research questions were examined through having 35 EHOs assess two kitchen environments on the LSE campus. Each kitchen appeared to function as a 'normal' kitchen (e.g. chefs and kitchen assistants present, food being prepared, equipment in use), but were designed so that a number of positive and negative information cues could be identified by the EHOs. These were expected to influence the ratings of the kitchen environments and participants completed a catering facility inspection report for each kitchen (the data for which was the 'dependent variable' in this study). In order to collect demographic and qualitative data, participants also completed a post-study questionnaire after each kitchen was assessed. The sample recruitment process, experimental scenarios, and post-study questionnaire are outlined below.

3.1. Participant recruitment

Participants were recruited as volunteers through an UK-wide mailing list of Local Authorities, provided by the Food Standards Agency. Researchers contacted local authorities and asked them to nominate one or two of their EHOs for a "research study on consistency". Potential participants were informed they would be reimbursed for any travel expenses and would be rewarded by a Continued Professional Development certificate for their participation (provided by the Chartered Institute of Environmental Health). To minimise any differences in EHO judgements due to preparation, participants did not receive any further information on the study before they participated.

3.3. Procedure

Each participant was booked in for a 4-hour time slot for the study at the LSE campus. When participants arrived on campus, they were met by one of the researchers. The researcher would then sit down with the participant to explain that the study would be based on 2 inspections of 30 minutes each, taking place in two separate kitchens on campus. Participants were informed that the study aimed to capture their judgements in a realistic and natural scenario. They were also told that their individual judgements and data would not be directly compared to those of others, as part of the researcher's commitment to anonymity and confidentiality. The participants then signed an informed consent form (see Appendix).

After signing the informed consent form, participants were given time to read the instructions for an inspection in the first experimental scenario (described below). After reading the instructions, the EHO was taken to the kitchen for a 30-minute inspection. After the inspection, the EHO was taken to a quiet place to complete an inspection form and a short questionnaire. This was followed by a 45-minute break. After the break, EHOs completed a second 30-minute inspection followed by another questionnaire, which was comprised of questions from the previous questionnaire. At this point, the EHO had finished their participation and handed in their materials.

Please see Appendix C for photos that illustrate the inspections in practice.

3.4. Experimental scenarios

Participants completed two 30-minute assessments of catering facilities at the LSE. The catering facilities were 'live' and staffed with real catering staff (a chef and a kitchen assistant), and EHOs were asked to use the standard FSA scoring system to assess the facility. Chefs and kitchen assistants were asked to behave normally (i.e. as they would do during a typical EHO inspection) in order to ensure consistency in their behaviour (i.e. to ensure they were not having to 'act' in an artificial way in the 35 scenarios in which they participated). As a broad observation, the two chefs differed somewhat in their personal style. The chef for scenario 1 was more experienced, and was concerned about ensuring the EHOs had a good impression of the facility and scenario (even though it was an experiment). The second chef had a more laissez faire style, was less experienced and concerned about the impression made to the EHOs. To make the scenario as realistic as possible, none of the researchers was present in the kitchen during the inspections. As the scenario presented the chef as employed by a "small catering company", the EHOs were not able to speak to the chef's manager. The chef was presented as the person responsible for the running of the kitchen and functioned as the EHOs' point of contact during the inspection.

The two kitchen scenarios were populated with props as to make them as realistic as possible. This included food, kitchen utensils, operational equipment (probe thermometers, fridges, ovens), cleaning materials (hand soap, sanitiser, cloths, mop) and training certificates. Furthermore, the chef in each kitchen was in possession of documentation for the operation

of the kitchen (the small caterers' "Safer Food, Better Business" (SFBB) guide, which was partially completed), a cleaning schedule, delivery records and temperature records for food and fridges. In addition, the EHOs were told to take for granted that they had seen 6 months' of historical records, including training records for the staff. The amount of 'scene-setting' for both scenarios was decided in consultation with an employee of the Food Standards Agency, who had several years of work experience as an EHO.

Scenario 1

EHOs examined a facility where the kitchen staff (a chef/manager and an assistant) were awaiting the delivery of pies for cooking in the oven, and the purpose of the kitchen was to serve pies to students in the student union. This scenario was designed to invoke an 'anchoring effect'. To achieve this, half of EHOs were given a negative previous report of the facility indicating it to have been assessed as a 2* quality facility at the last inspection (group 1). The other half of EHOs were given a positive previous report of the facility indicating it to have been assessed as a 5* quality facility at the last inspection (group 2).

Scenario 2

EHOs examined where the kitchen staff (a chef/manager and an assistant) were actively making sandwiches, and these were to supply functions around the university. This scenario was designed to invoke a 'confirmation bias effect'. To achieve this, on arriving at the premises half of EHOs were implicitly primed to view the premises positively. This was done through having a sign on a fridge (which was opposite the door through which they entered the facility) which indicated a negative cue (out of date records for food storage) about the standards of the catering facility (group 1). For the other half of EHOs, the sign contained a positive cue (e.g. the separation of raw meat from other products as indicated by a clearly labelled fridge) about the standards of the catering facility (group 2). In order to keep the information for each scenario consistent, the alternative sign not immediately cued (positive or negative) was placed elsewhere in the kitchen in a non-obvious location. EHOs were told this was a first inspection and therefore no previous ratings or other information about the kitchen was available.

Dependent variable

Before entering the scenarios, EHOs were given an inspection form and a notepad. EHOs were told they could use both to write on, but that their final assessment would have to be

recorded on the inspection form. Participants were given ample time to complete this form after each inspection, which culminated in their assessments of structure, hygiene, confidence in management (CIM), and overall food hygiene rating for the kitchen just assessed. The form, reprinted in the Appendix, was based on a template used by the Food Standards Agency, and captured data consistent with real inspections (i.e. qualitative data). This was developed with an employee of the Food Standards Agency with several years of work experience as an EHO, who provided guidance on the design of the simulated kitchens, the anchoring and confirmation biases, and also the questionnaire and documentation for recording the outcomes of hygiene inspections. Note that, due to time limitations, participants were not given a copy of the Food Law Code of Practice or the FSA's Brand Standard documents.

3.4. Feedback form and questionnaire

After each scenario was completed and participants had recorded their judgement on the inspection form, they completed a feedback form and a post-study survey. The feedback form allowed the EHOs to indicate how they would have spent their time if the inspection had taken longer than 30 minutes and how they would 'follow up' the inspection. Specifically, EHOs would indicate the feedback they would have provided to the chef after the inspections, what issues they would have raised and any other measures they would have taken (e.g. write enforcement letters). After completing the feedback form, participants completed a questionnaire that contained items on the preceding inspection, knowledge, attitudes and participant demographics.

<u>Preceding inspection items</u>

Participants were asked to indicate their confidence in the rating they had just given, as well as confidence in their own abilities (both on a 1-to-5 scale). They were also asked to provide three observations that had most influenced their judgement in the inspection, as well as the single most important factor in their assessment. Futhermore, they were asked how they thought their judgement compared to that of colleagues ("What percentage of your colleagues would have given the same rating as you?") and why. Finally, EHOs completed an affective scale after each scenario, a measure of their emotional state. The purpose of this scale was to examine whether affect influenced decision-making. Research shows that affect (e.g.

apprehension, tension, worry, happiness) influences judgement (i.e. where we have negative affect, we tend to make negative decisions), and the Positive and Negative Affect Scale (PANAS) was used to measure this (Lyubomirsky, King, & Diener, 2005; Watson, Clarke, & Tellegen, 1988).

Demographic factors

Participants were asked to report their gender, age, and years of experience. They also reported whether they had been on a training course recently. Finally, they completed a conscientiousness personality scale (see Lim & Ployhart., 2006).

4. Results

4.1.1 Descriptive data: Participants

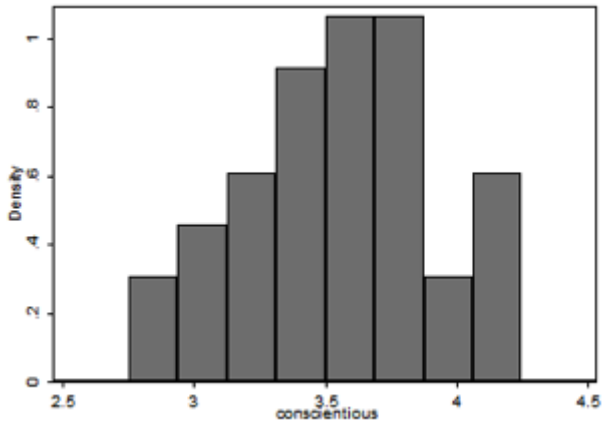
Table 1: Sample demographics.

Variable	Sample frequency or mean value
Gender	68.6% female 31.4% male
Age	45.0 years (St. Dev = 7.1, Min. = 30, Max. = 58)
Years of Experience	8.6% 1-5 years 28.6% 5-10 years 62.8% 10+ years
Last training course	85.7% last 6 months 14.3% last 12 months
Education	14.3 unspecified degree in Environmental Health or related 8.6% Diploma in Environmental Health or related 48.6% BSc in Environmental Health or related 28.6% MSc in Environmental Health or related

In total, 35 participants took part in the study. Table 1 reports the following demographic details of the sample: their gender; age; years of experience; when they last went on a training course and their formal qualifications. These data show that our participant sample is highly educated, trained and experienced. In these respects, the sample may thus be less representative of the population of EHOs as a whole; it is more reflective of the more experienced cohort in the population.

With respect to personality measures, Figure 1 reports the distribution of the participants' scores on the conscientiousness scale. Although the scores on the scale are relatively centered (with a tendency towards more conscientiousness), there is sufficient variance in the participant sample to act as an explanatory variable of consistency ratings.

Figure 1: Personality measure Histogram (Conscientiousness scale).



4.1.2. Descriptive data: Ratings

Each of the participants was asked to provide a food hygiene rating in both scenarios. One participant provided a rating of "3 or 4" for one of the scenarios. For the purposes of the statistical analyses that follow, this data point was given a value of 3.5. Figure 2 shows a histogram of the Food Hygiene ratings of either scenario. As the figure shows, there is some variation of ratings between scenarios: scenario 1 seems to receive higher ratings, on average. As shown in Table 2a, the average rating for scenario 1 is indeed higher. The variance in the scenario 1 (as expressed by the standard deviation) is lower than scenario 2. Note that tables 2b and 2c show the scores in the more conventional professional format of scores tables.

Figure 2: Histogram of Food Hygiene ratings across scenarios.

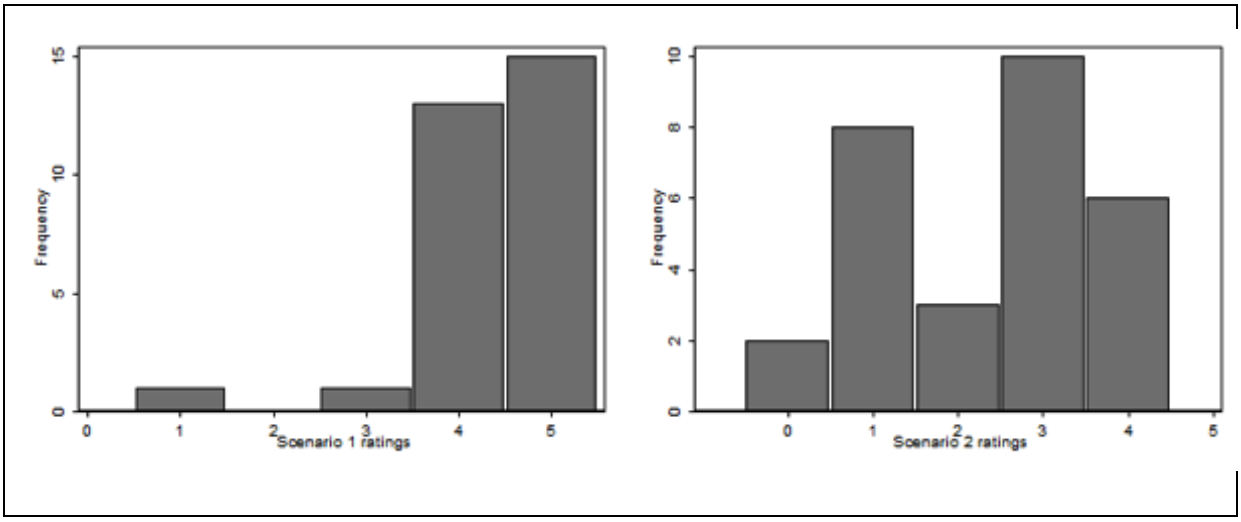


Table 2a: Food Hygiene ratings and components across scenarios.

Mean Values	Scenario 1	Scenario 2
FH Rating	4.4 (SD = 0.9, Min. = 1, Max. = 5)	2.4 (SD = 1.3, Min. = 0, Max. = 4)
Hygiene	4.3 (SD = 3.3, Min. = 0, Max. = 15)	10.6 (SD = 3.4, Min. = 5, Max = 20)
Structure	3.4 (SD = 2.7, Min. = 0, Max. = 10)	4.6 (SD = 3.9, Min. = 0, Max. = 10)
CIM	9.7 (SD = 4.8, Min. = 5, Max. = 20)	15.6 (SD = 5.6, Min = 10, Max = 30)

Table 2b: Scenario 1 Scores

Scores	0	5	10	15	20	25	30
Hygiene	9	23	2	1			
%	25.7	65.7	5.7	2.6			
Structure	12	22	1				
%	34.3	62.9	2.6				
CIM		13	17		5		
%		37.1	48.6		14.3		
FHR	0	1	2	3	4	5	
Rating		3		3	13	15	
%		8.6		8.6	37.2	42.9	

Table 2c: Scenario 2 Scores

Scores	0	5	10	15	20	25	30
Hygiene		5	22	7	1		
%		14.3	62.9	20	2.9		
Structure	12	14	9				
%	34.3	40	25.7				
CIM			17		17		1
%			48.6		48.6		2.9
FHR	0	1	2	3	4	5	
Rating		13	3	10	7		
%		37.2	8.6	28.6	20		

4.2. Research questions

Below we report the findings for each of the research questions outlined in section 2.

RQ1. To what extent are EHOs consistent in the assessments they make of catering outlets?

To examine whether EHOs were consistent in the assessments they make of catering outlets, we examined the distribution and standard deviations of ratings given by EHOs to the two experimental kitchens. As shown in Figure 1 above, there was substantial variance in ratings, especially in the second scenario. Using an ordered probability model, we examined to what extent the differences in Food Hygiene ratings can be attributed to scores in the categories of Hygiene, Structure, and CIM. An ordered probability model estimates the likelihood that a one-unit increase in an independent variable will result in an increase in an ordinal ranked dependent variable (a scale or rating system). It is necessary to use such techniques because ratings on a scale cannot be compared as regular quantities (the difference between a 1* and 3* rating is not the same as that between a 3* and 5* rating). Ordered probability models allow us to explain whether a change in one variable will lead to a change in Food Hygiene Ratings or its components. The regression results, shown in Table A1, regress the scenario Food Hygiene Ratings on its three components. The regression results show the scenario-specific results alongside combined results.

In scenario 1, we cannot detect a significant relationship between EHOs' impressions of the kitchen and their score. This is likely due to the lower variance between EHOs in this scenario. In scenario 2, we detect a significant and negative relationship between CIM and the final rating. This finding suggests that the higher variance in ratings in this scenario is mainly driven by how EHOs score confidence in management. Finally, the pooled regression results suggest that both Hygiene and CIM were influential factors in determining the rating. These results may reflect some of the between-scenario variance in EHOs' ratings, suggesting that considerations of hygiene and confidence in management are the main source of this variance.

To test for the influence of individual differences upon inspections, a series of multiple regression analyses were performed on the scores for Hygiene, Structure, and CIM, and the final rating attributed to the catering facility. Specifically, the predictive influence upon assessments of the variables "Years of experience", "Conscientiousness", "Age", and

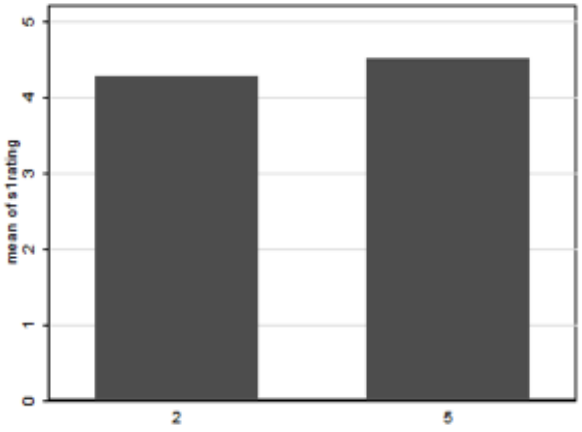
"Gender" was tested. The results of the regression analyses are reported in Tables A2 and A3. The first set of results (Table A2) show that experience is the only factor that consistently correlates with final ratings: more experienced EHOs give out lower ratings, particularly in scenario 1. The second set of regression results (Table A3) provide detail on how components of the rating are correlated with individual EHO differences. The coefficient estimates show that females provide significantly lower scores (better ratings) on structure across both scenarios, and a marginal effect of experience on hygiene – more experienced EHOs provided higher scores (worse ratings). This final result is in line with the correlations reported between experience and final ratings in Table A2.

RQ2a. Does 'anchoring' influence food hygiene ratings?

To examine whether the assessments of the experimental catering facilities by EHOs were influenced by previous inspections ('anchoring'), we compared the safety assessments made by the two experimental groups for experimental scenario 1. Group 1 was given a 'negative' (2*) previous inspection report, and group 2 was given a 'positive' (5*) previous inspection report. A series of rank-based pair-wise statistical comparisons tested for differences between the two groups. Rank-based comparisons allow us to test for significant differences in ratings between the two groups. The tests compare scores attributed to the categories of hygiene, structure, and CIM, alongside the final rating attributed to the catering facility. Figure 3a shows the average Food Hygiene rating for both groups and Figures 3b-d shows the average scores on Hygiene, Structure and CIM.

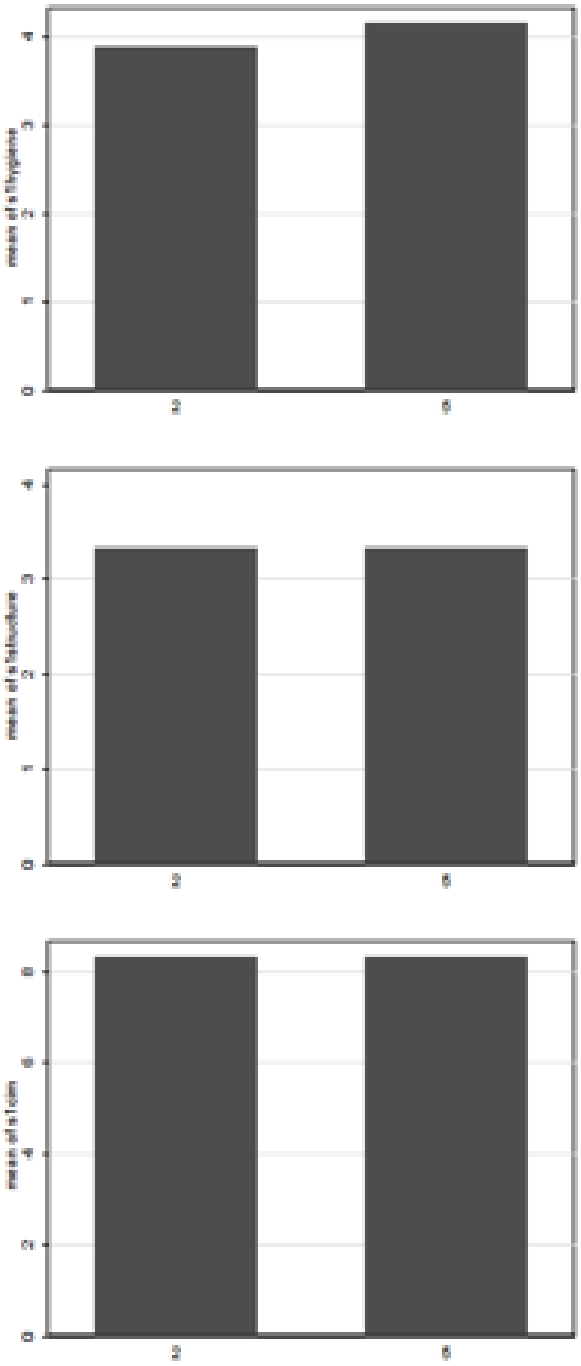
Pair-wise tests (two-tailed Wilcoxon rank-sum tests, all $p>0.1$) indicate no significant difference between the final Food Hygiene ratings in either group, as well as no difference between groups for any of the components (Hygiene, Structure and CIM). The data thus provide no evidence to reject the null hypothesis that the EHOs were not influenced by the anchor provided. We thus find no evidence for an anchoring effect on the previous rating.

Figure 3a: Average Food Hygiene ratings in scenario 1, by Group.



To test whether the treatment effect (the experimental manipulation of anchoring) might have been mediated by the individual differences of EHOs, a series of hierarchical multiple regression analyses were performed on the scores for hygiene, structure, and CIM, and the final rating attributed to the catering facility. Specifically, the predictive influence upon assessments of the variables "Years of experience", "Conscientiousness", "Knowledge", "Age", and "Gender" was tested. The results of the regression analyses are reported in Table A4 and show various patterns, most notably a gender effect. In line with the results from the non-parametric tests, we still find no effect of the anchoring manipulation (coefficient *group2*).

Figures 3b-d: Average scores (Hygiene, Structure, CIM) in scenario 1, by Group.



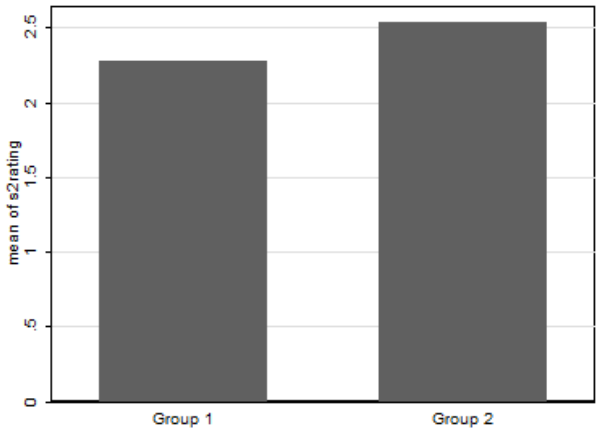
Notes: Panel 1 shows average Hygiene score; panel 2 shows Structure score; panel 3 shows CIM score.

RQ2b. Does 'confirmation bias' influence food hygiene ratings?

To examine whether the assessments of the experimental catering facilities by EHOs were influenced by previous 'confirmation bias', we compared the safety assessments made by the two experimental groups for experimental scenario two. Group 1 were cued to notice a

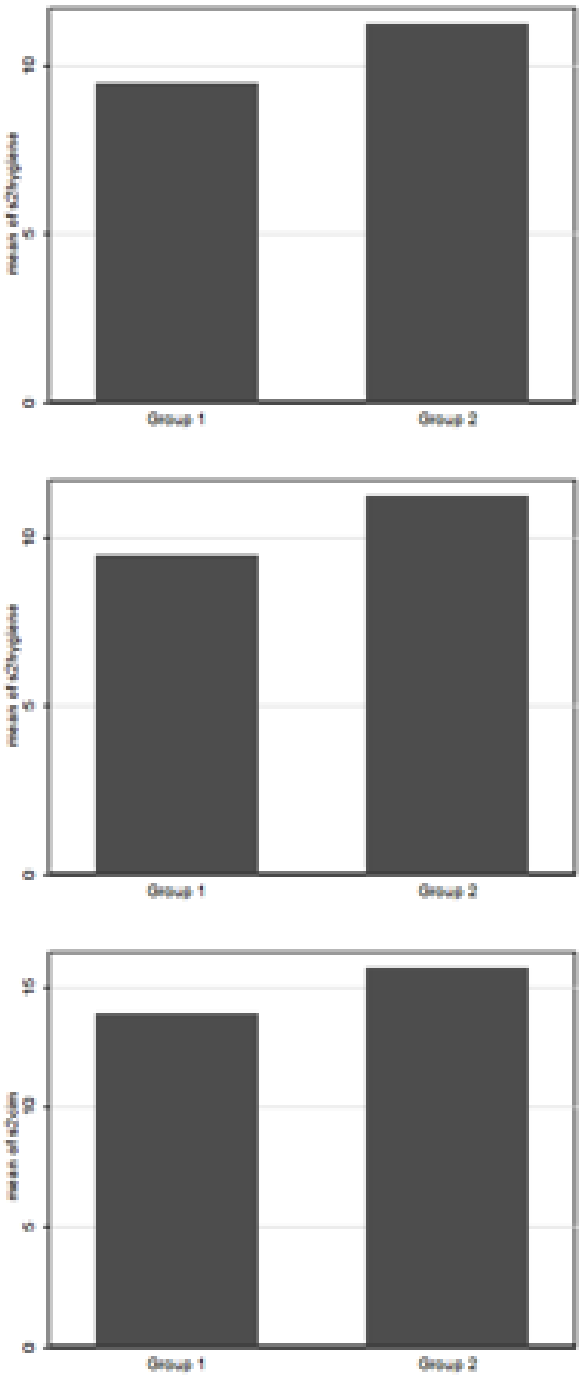
negative piece of information at the beginning of their inspection (which was intended to elicit a negative assessment of the facility), and group 2 were cued to notice a positive piece of information at the beginning of their inspection. A series of rank-based pair-wise statistical comparisons tested for differences between the two groups. Rank-based comparisons allow us to test for significant differences in ratings between the two groups. The tests compare scores attributed to the categories of hygiene, structure, and CIM, alongside the final rating attributed to the catering facility. Figure 4a shows the average Food Hygiene rating for both groups and Figures 4b-d shows the average scores on Hygiene, Structure and CIM.

Figure 4a: Average Food Hygiene ratings in scenario 2, by Group.



Pair-wise tests (two-tailed Wilcoxon rank-sum tests, all $p>0.1$) indicate no significant difference between the final Food Hygiene ratings in either group, as well as no difference between groups for any of the components (Hygiene, Structure and CIM). Although there are slight differences in both ratings and scores in the direction predicted by confirmation bias, these differences are not significant. The data thus do not show that EHOs were influenced by the order in which positive and negative cues were presented. We thus find no evidence for an effect of confirmation bias due to the order in which information was presented to EHOs.

Figures 4b-d: Average scores (Hygiene, Structure, CIM) in scenario 2, by Group.



Notes: Panel 1 shows average Hygiene score; panel 2 shows Structure score; panel 3 shows CIM score.

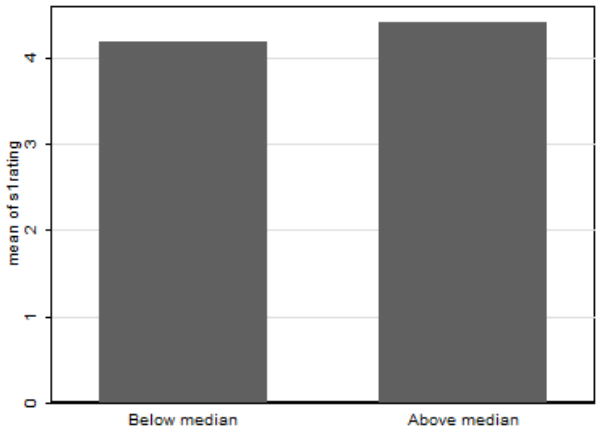
To test whether the experimental manipulation of confirmation bias was mediated by the individual differences of EHOs, a series of hierarchical multiple regression analyses were performed on the scores for hygiene, structure, and CIM, and the overall score attributed to the catering facility. Specifically, the predictive influence upon assessments of the variables

"Years of experience", "Conscientiousness", "Age", and "Gender" was tested. The results of the regression analyses are reported in Table A5 and show various patterns, although none of these seem to be consistently impacting scores apart from the interaction term between the experimental manipulation and EHO age. In line with the results from the non-parametric tests, we still find no effect of the confirmation bias manipulation (coefficient *group2*).

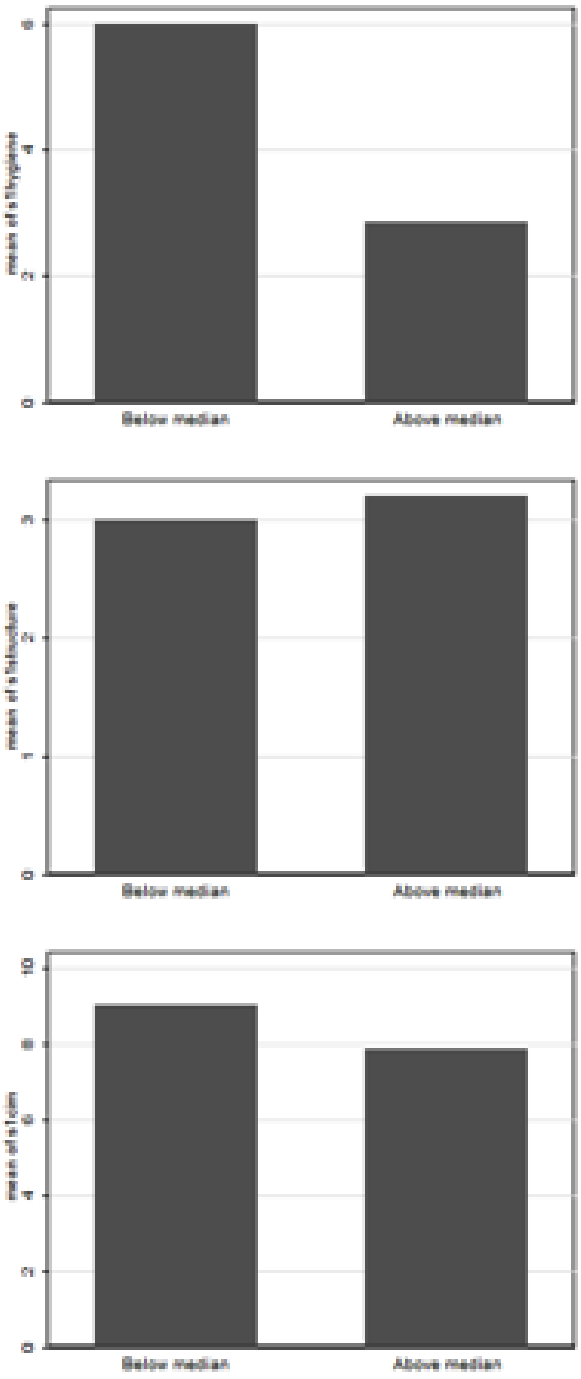
RQ2c. Does 'over-confidence' influence food hygiene ratings?

To examine whether the assessments of the experimental catering facilities by EHOs were influenced by 'over-confidence', we examined responses to the question item "Overall, how confident are you in the rating of the premises just examined" which all EHOs answered after each inspection. We performed a median-split on answers to this item (by each scenario), and thereby identified a 'high confidence' group and a 'low confidence' group. A series of rank-based pair-wise statistical comparisons tested for differences between the two groups. The tests compare scores attributed to the categories of hygiene, structure, and CIM, alongside the final rating attributed to the catering facility. Figures 5a and 6a show the average Food Hygiene rating for both groups and Figures 5b-d and 6b-d show the average scores on Hygiene, Structure and CIM. Although these figures show interesting patterns, none of the observed differences is statistically significant in the pair-wise tests (two-tailed Wilcoxon rank-sum tests, all $p>0.1$).

Figure 5a: Average Food Hygiene ratings in scenario 1, by Confidence.

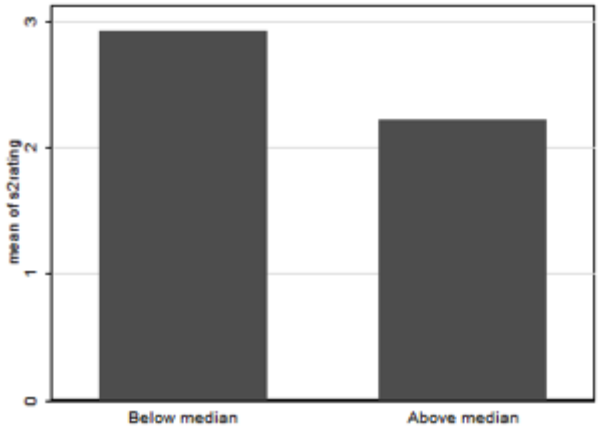


Figures 5b-d: Average scores (Hygiene, Structure, CIM) in scenario 1, by Confidence.



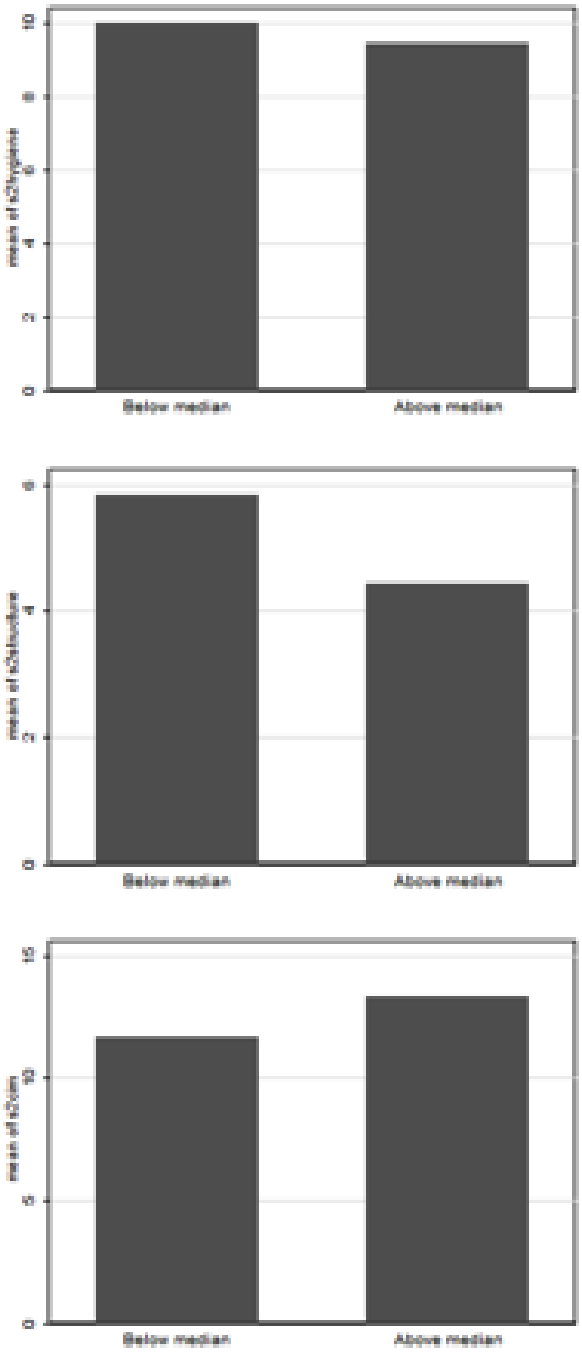
Notes: Panel 1 shows average Hygiene score; panel 2 shows Structure score; panel 3 shows CIM score.

Figure 6a: Average Food Hygiene ratings in scenario 2, by Confidence.



To take into account that over-confidence might be influenced by the different experimental scenarios and conditions, we then estimated two sets of multiple regression models in order to control for the impact of i) experimental scenario and confidence (treated as a linear variable) upon inspections. In addition, we also tested for the comparative influence of the other individual differences ("Years of experience", "Conscientiousness", "Age", and "Gender") upon the scores of EHOs for hygiene, structure, and CIM, and the overall score attributed to the catering facility. The results of the regression analyses are reported in Tables A6 and A7, and show various interactions between EHOs' self-reported confidence and their assessments of the catering facility. Most notable is the consistent interaction between gender and the confidence measure, with female EHOs' Food Hygiene rating increasing in self-reported confidence in both scenarios, and the CIM scores decreasing (better scores) in self-reported confidence as well. There thus seems to be an positive relationship between reported confidence after the inspection and the confidence in management score. For female participants, higher confidence in management scores (and higher final ratings) are correlated with higher confidence of EHOs' in their own judgements. We did not find this pattern in the judgements of male participants.

Figures 6b-d: Average scores (Hygiene, Structure, CIM) in scenario 2, by Confidence.



Notes: Panel 1 shows average Hygiene score; panel 2 shows Structure score; panel 3 shows CIM score.

5. Supplementary analysis

A rich volume of qualitative data was gathered through the administration of questionnaires, and can inform us about the EHOs' decision-making processes and the weight they give to different factors. Notably, in the post-scenario feedback questionnaire, participants were asked to report the three key observations that most influenced their rating for the premises they had just examined, and to identify the single most important factor.

5.1 Cues that influenced EHOs' decision-making process

Participants were asked to report the key elements influencing their ratings. Specifically, they had to list the three most important observations… To inspect these, the comments used EHOs to rate the scenario were thematically analysed. First, specific comments for each scenario were given generic issue-lables (by a single coder) in order that they could be aggregated together. Categories were based on the scoring notes in the FSA's Brand Standard. For example, a statement on "Attitude and disorganisation of the chef and lack of communication between him and event contractor" was categorised as "Poor chef attitude/competence". A comment on "Muddled, incomplete and inaccurate SFBB/documentation" was coded as poor procedures. All comments were categorised. Second, the number of issue-labels relating to each topic (e.g. Poor chef attitude/competence) across the entire dataset were aggregated (i.e. so the total number of comments relating to a specific issue on each scenario could be calculated). Third, issue-labels were divided into 'positive' and 'negative', so that 'positive' indicators (e.g. good chef knowledge) and 'negative' (e.g. poor documentation) indicators of kitchen management could be identified.

In total, across both scenarios, participants identified 13 positive indicators (cited 80 times in total): Kitchen in good repair/clean; Good chef knowledge; Good documentation; Good procedures/systems; Only preparing basic food; Signs of recent investment; Only preparing basic food; Kitchen in good repair/clean; Good chef knowledge; Good procedures/systems; Good PPE; Minimal food production; No non-compliances.

In total, across both scenarios, participants identified 27 negative indicators (cited 121 times in total): Poor documentation; Poor procedures/systems; Poor chef knowledge; Lack of observed food handling; Food handler cleanliness; Lack of management support; Broken equipment; Lack of hot food monitoring; Lack of space; Lack of training principles; Poor

previous inspection; Temperature control; Poor procedures/systems; Poor sanitation materials; Poor chef attitude/competence; Poor meat storage/contamination; Poor documentation; Poor sanitation practices witnessed; Lapsed records; New business; Poor training/supervision; Ad-hoc kitchen; Agency staff not competent; Broken equipment; Poor waste handling; Temperature control.

Results indicated that participants reported similar key elements in both scenarios. This is interesting because while scenario 1 generated similar ratings, scenario 2 produced more scattered ratings. The kind of observations and factors mentioned explain a lot about the decisions made by the EHOs. We will now explore the most-mentioned types of observations (Table 3) and factors (Table 4).

Table 3. Key observations for decision-making: Number of mentions.

Scenario 1		**Scenario 2**	
The supervision and training of the staff and the chef	23	The supervision and training of the staff and the chef	18
Food hygiene and structural requirements	20	FSMS	18
Documented procedures	20	Hygiene and practices witnessed	17
FSMS	14	Documented procedures	11
Nature of the food	8	Control measures in place to prevent cross-contamination	8

Table 4. Most important influence on Decisions: Most often mentioned.

	Scenario 1	**Scenario 2**
1	The supervision and training of the staff and the chef and FSMS	FSMS
2	Documented procedures	Hygiene and practices witnessed
3	Food hygiene and structural requirements	Food hygiene and structural requirements

Scenario 1

In scenario 1, most participants reported that food hygiene and structural requirements were crucial factors in their decision. Observations such as

- "Clean structure and equipment";
- "The structure and cleanliness were to a high standard";
- "The premises was clean and tidy"

were reported. The supervision, instruction and training of the staff and chef was also reported as very important and has a significant value for many participants, as these comments show:

- "The chef was open and confident";
- "The chef knew how to carry out practices correctly";
- "The chef was trained and knowledgeable";
- "The chef was very knowledgeable about food hygiene practices".

Only a few participants weren't satisfied with supervision and training of the staff and the chef but it was mainly about specific details such as pathogens or the lack of knowledge regarding the plans to expand.

Participants also reported that their rating decisions were influenced by the Food Safety Management System (FSMS), but disagreed on its completion and implementation. Several participants praised the good documented procedures:

- "Duty monitoring was being carried out";
- "The FSMS is in place";
- "Daily records support the FSMS";
- "The documentation is available on site with the records";
- "There are food safety procedures at critical points";
- "Quality of the SFBB pack";
- "Up to date records";
- "Documents were completed accurately".

On the other hand, an almost equal number of participants deplored the lack of dutifully completed safety procedures:

- "Lack of documentation and due diligence in keeping with a formalised food safety management system";
- "The SFBB requires more details";
- "The documentation does not fully relate to the operation";
- "The SFBB pack needs completion".

While the FSMS and documented procedures were among the most important factors for decision-making, participants gave different weights to those factors. The level of completion a document has to achieve to be considered "good" can explain this difference. Furthermore, participants who reported good documented procedures mentioned the quality of the FSMS, while critics emphasized the problems of the SFBB.

Scenario 2

In scenario 2, participants reported that their rating was principally influenced by their opinion of FSMS and documented procedures, hygiene and practices witnessed (i.e. food preparation, cooking, re-heating, cooling and storage of food), the supervision and training of the staff and the chef, as well as the control measures in place to prevent cross-contamination. The FSMS and documented procedures were mentioned most frequently:

- “No FSMS or records”;
- “No paperwork”;
- “Incomplete food safety records”;
- “Incomplete SFBB and lack of documented procedures”;
- “No dates on the training certificates”;
- “Muddled and inaccurate SFBB”.

But perhaps even more important, the supervision and training of the staff and the chef were often mentioned in a negative sense:

- “The chef was unaware that some fridges were not working”;
- “The chef is unaware of documented procedures”;
- “The chef appeared not to be bothered”;
- “Chef not able to understand SFBB”;
- “Inability to provide clear knowledgeable answers to the questions asked”;
- “Lack of understanding with high risk foods”.

Finally, hygiene and practices witnessed was also a major factor for their rating:

- “Poor practices witnessed”;
- “No hand soap to ensure proper handwashing”;
- “The catering assistant did not use the handwash basin”;
- “Issues in relation to E.Coli”;
- “Lack of disinfection of sink”;
- “Cross-contamination risks”.

Given that the participants were quite unanimous as to the criteria influencing their decision, the difference in ratings could be explained by the weight they gave to these criteria. We investigate this prediction statistically in the next section.

When we analyse the factors that the participants reported as most important for their decision-making in Table 4, we observe an interesting pattern. In scenario 1, the first most important element is the supervision and training of the staff and the chef, followed closely by FSMS, which is consistent with the number of times those facts are reported. However,

while being one of the most reported factors in the ratings rationale of the EHOs, the food hygiene and structural requirements came only third when participants were asked to rank the factors influencing their decision. The second factor in this hierarchy is the documented procedures. This can reveal that participants attach greater importance to the documented procedures – or lack thereof – than to the food hygiene and structural requirements. In scenario 2, the most important element is the FSMS. The second is the hygiene and practices witnessed; the food hygiene and structural requirements is the third most important, as in scenario 1. While the supervision and training of the staff and the chef is an oft-reported factor (see Table 3), participants indicate that such considerations are lower down the hierarchy.

5.2 Detailed justifications of the ratings

Scenario 1

In addition to the scores the EHOs gave to Hygiene, Structure, Confidence in Management and final rating, they were also asked to provide a rationale for their decision. In scenario 1, the data indicates that their rationale is rather similar to the three most important observations the EHOs mentioned in the previous section. However, there are a few differences (Table 5). While food hygiene and structural requirements and supervision and training of the staff and the chef were often mentioned as the three most important type of observations influencing their decision (Table 3), the documented procedures seemed to have more significant weight and came first in their justification for the ratings, as these remarks show:

- "Incomplete SFBB";
- "Some issues with the pack";
- "Good record keeping but it does not always reflect what is currently going on on site";
- "SFBB requires more development".

Again in scenario 1, participants had previously reported that they had been greatly influenced by the supervision and training of the staff and the chef. However, hygiene and practices witnessed were mentioned more frequently in their ratings' rationale:

- "No raw food so low risk";
- "Low risk because pre-cooked pies";
- "Minimal cross-contamination risk";
- "No hygiene non-compliance found";
- "Whilst boards and knives block can pose cross-contamination risk, there is no raw meat at present".

The fact that the EHOs justified their ratings mainly on the basis of the documented procedures, hygiene and practices witnessed can be explained by the time of the administration of the different questionnaires. They had to explain their rationale on the inspection form. But the question about the three most important observations influencing their decision was asked in the post-scenario feedback questionnaire. At the time of the inspection, participants were particularly concerned about the completion of the documented procedures and hygiene and practices witnessed. When the EHOs had to report and rank the factors influencing their decision in a later questionnaire, the food hygiene and structural

requirements and the supervision and training of the staff and the chef where the most frequently mentioned factors. This reveals an interesting pattern: staff training and competence are recognised as important observations, but not often mentioned as strong influences or rationales. This may be due to the more subjective nature of staff competence and other CIM-related factors.

To investigate the extent to which we could relate certain patterns in EHOs' rationales to their rating, we also conducted a regression analysis of rating on rationale. As before, a single coder coded the occurrence of the most commonly cited factors (documented procedures, the FBO being a new business, supervision and training of the staff and the chef, hygiene and practices witnessed) in the rationale of each individual participant and investigated their relationship with the rating given. As before, categories for coding were based on the FSA's Brand Standards document.

Note that positive and negative indicators were coded separately; this to account for the different influence that these will have on Food Hygiene ratings. In scenario 1, we did not find that any of these mentions correlated significantly with the ratings (Tables A8a-8d). It does not appear, therefore, that any of these factors in isolation strongly affected Food Hygiene rating in scenario 1.

Scenario 2

In scenario 2, the three most important types of observations reported by EHOs were the FSMS, hygiene and the supervision and training of the staff and the chef. When asked to justify their reasoning behind their rating, participants also reported different elements. Similarly to scenario 1, the documented procedures was the most frequently mentioned factor in their rating's rationale:

- "No paperwork";
- "Lack of documents";
- "Incomplete paperwork";
- "SFBB muddled, incomplete and inaccurate";
- "Need to complete the SFBB sections that are relevant";
- "SFBB requires further development";
- "Some gaps in SFBB but making progress";
- "SFBB pack not fully reflective of the operation".

Hygiene and practices witnessed seemed to matter equally in the most important factors influencing their decision and in their rating's rationale:

- "Ham being left out – bad hygiene";
- "Lack of hygiene";
- "Need a greater understanding of the importance of the bin";
- "Cross-contamination issues";
- "Must ensure soap available by hand wash basin";
- "Risks of E.Coli and cross-contamination";
- "Some lapses in food hygiene".

Similarly, the supervision and training of the staff and the chef was mentioned both in the most important factors influencing their decision and in the rating's rationale:

- "The chef was not aware of policies";
- "The staff had good safety knowledge";
- "Not great deal of confidence in chef";
- "The chef had a lack of understanding in some areas of food hygiene".

When they justified their rating, participants seemed to value the documented procedures greatly, while in the most important reasons influencing their decision it was the FSMS that were regarded as most important. In both scenarios, documented procedures came first when the EHOs had to justify their rating during the inspection. In the post-scenario feedback form, the food hygiene and structural requirements and the FSMS were weightier. While in scenario 1 the supervision and training of the staff and the chef was less important in their rating's justification, this factor was equally important both during and after the inspection in scenario 2. We noted specifically that the predominant pattern in this scenario was participants' lack of trust in the chef's competence.

As for scenario 1, we conducted a regression analysis on the relationship between Food Hygiene rating and commonly mentioned factors in EHOs' rationale (new business, supervision and training of the staff and the chef, hygiene and practices witnessed). As before, we conduct a separate analysis on positive and negative mentions. In contrast to scenario 1, where we found no significant correlations, we find a significant correlation between negative mentions of supervision and training of the staff and the chef and the rating. That is, participants who negatively comment on the chef's competence in their justification

in scenario 2 give lower ratings. For the other commonly mentioned factors (including positive mentions of the chef), we do not find such correlations (Tables A9a-9d). Since this was only mentioned by 7 participants, this appears to be a particularly strong influence on final rating. Although factors relating to the supervision and training of staff are less frequently mentioned than other factors, they do seem to exert a powerful influence (especially when judged negatively). In line with our earlier results on the frequency of these factors in the three key observations, this suggests that the impression of the kitchen staff competence plays a role in final judgements that may not necessarily be picked up in a written justification for the score.

Table 5: Rationale for ratings.

	Scenario 1	**Scenario 2**
Documented procedures	17	24
Hygiene and practices witnessed	17	20
Supervision and training of the staff and the chef	11	7

5.3 Follow-up actions

Participants were asked to say what actions they would have taken for this establishment if they were employed by the responsible Local Authority. We now discuss the prevalent patterns in participants' answers, followed by a regression analysis of the relationship between these answers and Food Hygiene ratings.

In both scenarios, participants systematically mention that they will leave a report of inspection at the FBO, and will also send a letter. Also in both scenarios, participants report they would offer further advice or coaching:

- "Contact them on the phone to discuss legal responsibilities";
- "Advice to chef on SFBB";
- "Would offer assistance/advice to complete SFBB";

- “Would give further links to allergens resources, SFBB, training, DVD's, E.coli guidance”;
- “Would’ve offered a coaching visit to help them implement SFBB as the business was keen”;
- “Possible advisory visit if requested”;
- “SFBB coaching at a later date”;
- “Meeting offered with senior staff”;
- “Provide additional guidance/advice to chef and kitchen assistant on SFBB completion and implementation”;
- “Offer of coaching to improve food safety”;
- “I would have offered a coaching session on the SFBB implementation”;
- “Possible SFBB training for head chef”.

Willingness to offer advice thus seems to be strong across scenarios. But along other dimensions, the patterns of answers differed markedly. We find a large difference in the intention to revisit the premises. In scenario 1, only a quarter of the participants would revisit; and in scenario 2, more than half of the participants would. We now discuss some of the reasons behind those decisions.

Scenario 1

In scenario 1 the average rating is good, which would imply that a revisit is unnecessary. Participants who would conduct a further visit mentioned the following reasons:

- “To check the FSMS”;
- “Contact the FBO on the phone and revisit if not confident in what FBO said”;
- “Revisit to check SFBB pack”;
- “I would revisit to check is SFBB had been completed and if they had implemented the plans to start handling raw meat”;
- “I would consider revisiting once the on-site pie manufacturer had started and to check his SFBB”.

Documented procedures and future risks seem to be the main motives for a future inspection in scenario 1.

We also conducted a regression analysis that investigates the relationship between EHOs’ intention to follow-up their visit and their rating (Tables A10a-10b). We found no relationship between participants’ intention to provide coaching and the rating, but we found

a significant relationship between participants' intention to revisit and the rating. For lower ratings, EHOs were more likely to say they would revisit the FBO. Note that we also found a significant relationship between intentions to revisit and the CIM score (but not for any of the other rating components). This suggests that CIM is the main factor that determines EHOs' intentions to revisit.

Scenario 2

In scenario 2, half of participants say they would visit the premises again. Most participants do not specify why they would do a revisit, but those that do mention two main reasons. The first reason for a revisit has to do with the documented procedures:

- "Follow up visit to check paperwork";
- "If my workload allowed I may have carried out a revisit to check his paperwork was completed in full";
- "Revisit in 2 weeks to check the non compliances and SFBB";
- "Possible revisit to check progress on SFBB";
- "Revisit in 4 weeks to review the SFBB pack".

The second reason for a revisit is to make sure problems have been resolved:

- "Would follow up in a couple of weeks to ensure important issues had been addressed";
- "Re-visit in 2 weeks to ensure HACCP policies fully implemented";
- "Revisit because needs hygiene improvement – notice if no progress";
- "Revisit after coaching visit to check compliance".

As in scenario 1, we also conducted a regression analysis of intentions to follow-up in scenario 2 (Tables A11a-11b). We found the same results as in scenario 1: no relationship between rating and coaching but for lower ratings and poorer CIM scores, EHOs were more likely to say they would revisit the FBO. This is a positive signal, suggesting that EHOs offer further assistance to those FBOs that need it, rather than on the basis of other factors. Ratings are more strongly correlated with re-visits than offer to provide coaching, suggesting that the offer of coaching is more dependent on the EHO's individual characteristics and style, whilst re-visits are more clearly tied to outcomes.

5.4 Comparison with colleagues

In the post-scenario questionnaire, participants were asked to give reasons why other EHOs might have made a different assessment than them. Six main factors come up regularly (Table 6), and will be explained in detail in this section. These key points are (i) the documented procedures, (ii) confidence in management, (ii) hygiene and practices witnessed, (iv) the fact that the FBO is a new business, (vi) weighting of evidence and (vi) the questions asked.

Table 6: Reasons given by EHOs for why their colleagues might have given different ratings.

Differences in ratings	**Number of citations**
CIM score	22
Documented procedures	12
Hygiene and practices witnessed	9
Questions asked	8
Weighting	7
New business	3

Documentation

Several participants reported that their colleagues might have given more weight to the documented procedures:

"Greater emphasis on the lack of formal system than risk at time of visit";
"I think some may concentrate heavily on the presence or absence of paperwork";
"May have taken a harsher view at the failure to have completed the pack fully";
"SFBB not completed - more may view this more harshly";
"I suspect that some may have given a lower score when they saw the incomplete SFBB";
"Some may give worse score for not having adequate documentation".

Confidence in management (CIM)

Some participants assumed that their colleagues might have given a better CIM score:

- “Some officers may just see a very clean, structurally excellent kitchen and not consider the management system to be as important”;
- “Some EHO's may have given a better score for CIM”;
- “Some might have considered the management system to be effective enough”;
- “Might not consider importance of CIM section of rating”;
- “Some EHO's may have given a better score for CIM”.

Others thought their colleagues might have given a worse score to the confidence in management:

- “Some may have given worse score in CIM because no paperwork”;
- “CIM score - based on previous history - some colleagues may have given lower score”;
- “Possible worse score in CIM”;
- “Some colleagues may have scored CIM more harshly”.

And others mentioned that their colleagues might have judged the confidence in management differently:

- “I know the CIM score is the biggest source of lack of agreement”;
- “There seems to be the most problems around the management score”;
- “Different scores for CIM”;
- “Management score”.

Given the mean scores and variance we see in the two scenarios (see Table 3 in the previous section), the EHOs are correct: CIM is indeed the FH rating component with the highest variance.

Hygiene and practices witnessed

Some EHOs gave a good score for hygiene only on the basis of the current risks: “As no raw meat, eggs etcetera the cross-contamination issues may not have been viewed as seriously”. Others would be more concerned about future risks if the business would expand or handle raw meat:

- “Some EHO's may also not have considered the fact that the chef has plans to expand the process which would change the score dramatically”;
- “If they were concerned about future plans of the site”;
- “It was broadly compliant for this operation but if it were more complex it may not have been”.

New business

Some EHOs might be more lenient because the premises are a new business, thus they assume that the business would improve quickly:

- “New business so may have offered help and score better”;
- “Some may have given a 5 for management as the business has only been open 3 weeks”.

Some EHOs reported that they gave a better score in CIM in order to give a chance to comply. They also mentioned that they were confident that the chef had good training and that he would complete the documentation and implement the food safety procedures. They reported that the documented procedures should be given less weight since it is a new business. Some were also reluctant to give bad ratings because it can have a devastating impact on new premises.

But on the other hand, some EHOs would be less lenient: “Those in busy LAs may not have the time to nurture new businesses such as this”. A new business was reported as a source of lesser confidence. Furthermore, while some EHOs praised the brand new kitchen, others were suspicious and reported that new premises are usually dirtier when a second inspection is conducted. They were also less tolerant because they reported that a new business is easier to operate.

Different weighting

Many EHOs reported that their colleagues might have given different weights to different areas of inspection:

- “They may have gone into much more depth on some issues - time allowing - and weighted scoring differently as a result”;
- “Some focus on small details rather than making an overall assessment of the risks”;
- “Others may have weighted it differently”;

- “Focusing on other areas - giving more weight to some issues than others”;
- “EHOs may have concentrated on different areas of the inspection”.

The participants also mentioned the previous history as influencing their weighting. The general attitude of the FBO and the willingness to comply with the law were reported to be important because they are more difficult to correct than other factors. Some EHOs said that each inspection is different, and they can only judge what they see at the time of the inspection, which can impact their rating. Whether the premises are busy or not at the time of the inspection can change the weighting and other issues might also appear.

Questions asked

EHOs might have asked different questions and therefore received different answers, which could have produced different ratings:

- “Not asking the right questions at the chef (or any questions)”;
- “May have asked different questions”;
- “They may have questioned differently”;
- “They may have asked different questions or received different answers”;
- “May not have asked the same questions”.

The questions EHOs ask and the weighting they give to different areas might be related. Some participants reported experience to be linked to the kind of questions asked, and how their colleagues value the health risks as oppose to incomplete documentation. Other participants mentioned that they were trying to build a good relationship with the premises, thus influencing the questions asked.

Overall, what can create differences in ratings is the weight given to the documented procedures and the confidence in management. Those seem to be the main areas EHOs disagree on. Furthermore, some EHOs might only take into account the current health risks while others might give a different score because they would be concerned about future risks. Finally, there might be differences in the questions asked and the weight given to each area of inspection, thus resulting in different scores.

Confidence and accuracy

Finally, we tested whether participants correctly guessed the percentage of their colleagues that would have given the same rating as them. All participants were asked to state this

percentage in brackets of 20 percentage points (e.g. 0-20%, 20-40%..). In scenario 1, we find that EHOs who gave higher ratings were more confident that others would have done the same: those who give a 4* or 5* final rating are most likely to say that between 60-80% or 80-100% of their colleagues would have given the same rating (Table 7). The actual distribution of ratings suggests that most of these judgements are overconfident, however. Only a handful of EHOs correctly guesses the likelihood that another EHO would have given the same rating. In scenario 2, EHOs do not fare much better: the vast majority still overestimate the likelihood that others would have given the same rating (Table 8). Note that overconfidence does not seem to be related to the level of the rating, however. This robust finding suggests that, although EHOs are aware of differences between each others' ratings, they still underestimate inconsistency across ratings. This level of overconfidence could be a useful starting point for thinking about ways of reducing the variance in ratings, such as debiasing techniques where the decision-maker is asked to take "someone else's perspective".

Table 7: Scenario 1

Rating	"What percentage of your colleagues would have given the same rating as you?" (Correct answer for participant sample in highlighted cell)					
	0-20%	20-40%	40-60%	60-80%	80-100%	Actual ratings (n=35)
None			1			3%
1				2	1	9%
2						0%
3			1	1	1	9%
4		1	2	6	4	37%
5			2	5	8	43%

In our regression analysis, we found that there is a strong positive correlation between an EHO's confidence in his or her own judgement and the estimated percentage of colleagues that would have given the same rating (Tables A12 and A13). This pattern holds for both scenarios. Given the observed inconsistency between participants' judgements about colleagues and the real frequencies, this correlation is striking. One promising avenue of further research in this area is whether EHOs' judgements will vary less if some element of peer review or comparison is introduced.

Table 8: Scenario 2

Rating	"What percentage of your colleagues would have given the same rating as you?" (Correct answer for participant sample in highlighted cell)					
	0-20%	20-40%	40-60%	60-80%	80-100%	Actual ratings (n=34)
None				1	1	6%
1			2	7	4	38%
2				2	1	9%
3		1	3	5	1	29%
4			2	1	3	18%
5						0%

6. Discussion.

This report detailed the results of an experimental study on the consistency and variance of the assessments of Environmental Health Officers (EHOs), conducted at the London School of Economics and Political Science in March 2015. The experimental results suggest that EHOs are not susceptible to anchoring bias on previous ratings, nor to confirmation bias triggered by the order in which information was presented to them.

In our study, confidence in management (CIM) is the component with the strongest influence on the final ratings of EHOs. A large part in the variance of ratings between EHOs and experimental scenarios was explained by differences in CIM scores. For female EHOs, the results also suggest that confidence in their own judgement is related to their CIM scores. In our qualitative analysis of the data, we find that CIM factors feature in explanations of "key observations" but that they are rarely mentioned as justifications for a particular rating.

In reflecting upon the findings of the report, it is necessary to consider why inspectors were not found to be susceptible to anchoring bias nor confirmation bias.

First, it is possible that anchoring and conformation bias do not influence EHO food safety inspections. Conducting a food safety inspection is highly complex – for example it involves the evaluation of premises, interpretation and integration of information, conversations with staff, fact-finding, translation of qualitative observations into a quantitative scale, and developing recommendations. Decision-making is highly situated, and occurs over a long period of time (30 minutes in the experimental scenario, longer in reality). The majority of research into heuristics and biases has been conducted in experimental scenarios, often with participants conducting relatively simply problems (e.g. economic games). In comparison to these settings, the sheer complexity of the work done by EHOs may mean that the biases explored within the current study having little bearing on decision-making. Put simply, EHO decisions are based on an extended and wide-ranging evaluation of information (plus personal styles that influence how this occurs) and it might be unrealistic to expect one heuristic or bias (or one attempt to elicit these) to influence decision-making. Yet, this is not to say EHOs may not be biased in their decision-making. EHO assessments were found to vary considerably within the experimental scenarios (despite the scenarios being highly consistent across EHOs), indicating heuristics and biases to be shaping decision-making. Rather, EHOs showed a highly complex pattern of decision-making that differed from inspector to inspector, with individual EHOs putting different emphasis on the information

collected to inform inspection ratings (e.g. the significance placed upon staff competence or the wider organizational environment), and having different interpretations of how they should use that information (e.g. how it should influence final scores). We also note that the importance of CIM factors, such as the chef's competence, can have strong influences on ratings even though they are rarely mentioned as a justification or rationale. This is consistent with naturalistic decision making theory (Klein, 2008), which posits that decision-making is influenced by a combination of expertise, environmental constraints, and heuristics and biases. It may have been the case that, had these factors been mentioned as part of the prior information about the outlet, the effect of confirmation bias would have been more pronounced.

Second, and elaborating on the above, the sample used in the study may not have been a typical sample of EHOs. Participants were recruited at short notice and were highly enthusiastic. Many were highly experienced, and consistent with Klein's (2008) naturalistic decision-making approach, this is likely to have influenced how heuristics and biases influenced EHO inspection ratings. More specifically, expert decision-makers tend to make decisions based on previous experience ("pattern matching"), whereby they develop a personalised set of heuristics (routines and behaviours) for making a decision (e.g. questions for chefs, order of information evaluated), with decisions often being based on comparisons with previous experiences (e.g. what is normal, what decision was made in a similar previous experience?). Thus, EHOs are likely to develop their own personalised heuristics (e.g. information or behaviour that is deemed highly indicative of kitchen safety) for making food safety decisions. Whilst these are still likely to produce bias, they will be particular to individual EHOs, and will be a product of individual experience and expertise. This argument is borne out by the qualitative data, which shows a variety of cues (across EHOs) to have shaped decision-making. Whilst factors such as the knowledge and attitude of the chef were often identified as important, aspects such as documentation, procedures and systems, and previous ratings varied in their significance to EHOs. The pattern of personal heuristics is also borne out by participants' perception of how their rating compares to that of their colleagues: nearly all participants overestimate how likely others are to give the same rating to the FBO. The more confident the EHO is about his or her personal score, the greater the overestimation. Studying these patterns, perhaps coupled with encouraging EHOs to reflect on these patterns and share them with others is a promising area for further research and

training activity. This is especially true with regard to scoring CIM through quantifying staff training and competence factors.

Third, aspects of the experimental design may have limited the likelihood of eliciting or detecting bias in EHO decision-making. One, despite the scenarios attempting to recreate a real kitchen environment, their fidelity was not optimal. In the post-study feedback, participants highlighted some of the limitations in the scenarios (e.g. no wider management structures, limited use of the cooking facilities, new chef), and many pointed out that whilst the experiment was much better than traditional methods of training (e.g. computers, paper and pen) and quite realistic, their behaviour and ability to assess the facility was constrained. Furthermore, whilst we tried to ensure the experimental scenarios were highly consistent in design, identifying the 'actual' hygiene assessment score of the kitchen was difficult (i.e. so we could detect bias). The simulated kitchens were designed in conjunction with an EHO working for the FSA, yet calculating a 'correct answer' for the scenario was impossible, because there is a certain degree of subjectivity in applying the FSA rating scales to assess food safety. This means that our calculations about 'bias' are based on variations amongst EHOs in how they assessed the scenarios, and not variations from an objective standard (making it harder to ascertain 'bias'). It was also noticeable that some of the participants were very harsh in attributing ratings (e.g. 1* and 2*), and these very conservative scores, are quite unusual against the wider distribution of food safety ratings (where 1* and 2* are quite rare), and are indicative of the Hawthorne effect (where participants are highly conscious of being observed, and thus engage in socially desirable or conservative behaviours because they feel they are being tested). Furthermore, the study cues used to elicit bias in decision-making may not have been optimal. In particular, considering the wide range of information used to generate EHO safety ratings, the study cues (poor documentation, poor previous hygiene inspections) may have had little influence upon ratings. Whilst participants did report using these cues in making their assessments, other factors (in particular discussions with the chef) were far more prominent (negating the impact of the bias-eliciting cues). This is supported by many aspects of our dataset, indicating that confidence in management and the chef's knowledge were the most significant factors affecting EHOs' rating. Finally, whilst the experimental scenarios were consistent, the EHOs were varied in their experience and demographics. Whilst we did control for this in our analyses, previous experience of conducting inspections might have influenced decision-making in ways that were not able to anticipate.

Finally, it was noticeable that scoring in the first (which was generally rated well) scenario was more consistent than the second scenario (which was generally rated poorly). Without post-study interviews, it is not possible to categorise why this might be the case. However, qualitative data revealed that in the second scenario, poor scores were generally attributed by EHOs due to the poor performance of the chef (unlike the first scenario, where the chef was generally agreed to be competent. However, the degree to which EHOs took this into account varied. Some focussed more on the technical aspects of the kitchen (e.g. hygiene), and placed less weight on the chef. Others placed greater store on this. This might reflect aspects of experience (e.g. what is judged as 'normal', training, and inter-personal abilities (i.e. for EHOs interviewing the chef), and should be the subject of further study.

In reflecting on the results of the research, it is necessary to consider future avenues of research and training activity. In particular, we identify three possibilities.

First, future research may wish to focus more upon the patterns and cues that EHOs use to make food safety decisions, rather than the generic biases that may influence these. As discussed above, individual EHOs are likely to develop personalised strategies for assessing food safety and to place emphasise and importance upon different information sources (based on their previous experience). Determining these, and understanding how they may create bias and/or lead to effective decision-making, appears essential. In particular, such insight would be highly useful for developing training programmes, and for developing a more unified and shared set of standards for assessing kitchen safety. Particularly when it comes to more subjective judgements, such as choosing questions for assessing staff competence within a specific situation, it could be highly beneficial to capture, share and discuss these strategies within the wider community of EHOs. Our the study indicates that future research may wish to focus less on heuristics such as confirmation bias and anchoring, and more on representativeness (i.e. what is judged to be 'representative' of a good or poor kitchen, and the tell-tale signs of this). Our qualitative analysis of negative and positive influences of EHO assessments provides a starting point for doing this.

Second, future research may wish to leave the laboratory, and focus on heuristics and biases in the real world. This might involve use different methods (e.g. 1st person video recording of

inspections), or comparing EHO inspections for the same kitchen. Whilst the laboratory environment used in this study was highly appropriate for initially examining whether and how EHO decision-making is influenced by bias, field research is required to better understand decision-making. Having established considerable variance in EHO decision-making, future research needs to better establish why this occurs, how it can be understood, and if it requires some amelioration. Simply, the complexity of a real kitchen environment cannot be easily replicated, and decision-making is highly contextual.

Third, whilst the limitation of laboratory-style studies have been discussed, they may be more appropriate for inexperienced EHOs. The EHOs in our sample were highly experienced, and likely had behavioural patterns and norms that influenced decision-making (and minimised cues designed to invoke bias). This problem may be less likely for inexperienced EHOs, who will not have the same experiences and depth of knowledge influencing decision-making. Thus, to establish whether and how biases influence the decision-making of EHOs, a less experienced sample may better facilitate the exploration of this within a laboratory setting. An alternative, although this study strived to 'recreate' a functioning kitchen, our participants highlighted points where it was not 'high-fidelity' (e.g. no eating area, minimal documentation, unclear history of the establishment). Future research may wish to invest in creating a more in-depth and immersive experimental environment, and this would require expert input from the FSA.

REFERENCES.

Bazerman, M., & Moore, D. A. (2012). Judgment in managerial decision making. Hoboken, NJ: John Wiley Sons Inc.

Cannon-Bowers, J. A., & Salas, E. E. (1998). Making decisions under stress: Implications for individual and team training. American psychological association.

Croskerry, P. (2003). Cognitive forcing strategies in clinical decisionmaking. Annals of emergency medicine, 41, 110-120.

Croskerry, P. (2003). The importance of cognitive errors in diagnosis and strategies to minimize them. *Academic medicine*, *78*(8), 775-780.

Croskerry, P., & Norman, G. (2008). Overconfidence in clinical decision making. The American journal of medicine, 121(5), S24-S29.

Croskerry, P., Singhal, G., & Mamede, S. (2013). Cognitive debiasing 2: impediments to and strategies for change. BMJ Quality and Safety, 22, 65-72.

Gigerenzer, G., & Gaissmaier, W. (2011). Heuristic decision making. Annual review of psychology, 62, 451-482.

Janis, I. L. (1972). Victims of groupthink: a psychological study of foreign-policy decisions and fiascoes.

Johnson, E. J., & Weber, E. U. (2009). Mindful judgment and decision making. Annual review of psychology, 60, 53.

Kahneman, D., & Frederick, S. (2002). Representativeness revisited: Attribute substitution in intuitive judgment. In T. Gilovich, D. Griffin & D. Kahneman (Eds.), Heuristics and biases: The psychology of intuitive judgment (pp.49-81). New York : Cambridge University Press.

Kahneman, D. (2011). Thinking, fast and slow. New York: Macmillan.

Klein, G. A. (1993). A recognition-primed decision (RPD) model of rapid decision making (pp. 138-147). Norwood, NJ: Ablex Publishing Corporation.

Larrick, R.P. (2004) Debiasing (Chapter 16). In: Handbook of judgment and decision making. Malden: Blackwell Publishers.

LePine, J. A., Colquitt, J. A., & Erez, A. (2000). Adaptability to changing task contexts: Effects of general cognitive ability, conscientiousness, and openness to experience. Personnel Psychology, 53(3), 563-593.

LePine, J. A., & Van Dyne, L. (2001). Voice and cooperative behavior as contrasting forms of contextual performance: evidence of differential relationships with big five

personality characteristics and cognitive ability. Journal of Applied Psychology, 86(2), 326.

Lim, B. C., & Ployhart, R. E. (2006). Assessing the Convergent and Discriminant Validity of Goldberg's International Personality Item Pool A Multitrait-Multimethod Examination. Organizational Research Methods, 9(1), 29-54.

Lyubomirsky, S., King, L., & Diener, E. (2005). The benefits of frequent positive affect: does happiness lead to success?. Psychological bulletin,131(6), 803.

Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. Review of general psychology, 2(2), 175.

Oswald, M. E., & Grosjean, S. (2004). Confirmation bias (Chapter 4). In: Cognitive illusions: A handbook on fallacies and biases in thinking, judgement and memory. New York: Psychology Press.

Tetlock, P. (2005). Expert political judgment: How good is it? How can we know?. Princeton University Press.

Thaler, R.H. & Sunstein, C.R. (2008). Nudge: Improving Decisions about Health, Wealth, and Happiness. Yale University Press: New Haven, CT.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. Science, 185, 1124-1131.

Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. Science, 211(4481), 453-458.

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: the PANAS scales. Journal of personality and social psychology, 54(6), 1063.

Zsambok, C. E., & Klein, G. (Eds.). (2014). Naturalistic decision making. New York: Psychology Press.

APPENDICES

Appendix A: Statistical analyses

Table A1: Ordered Probit regression model, determinants of FH Rating.

	Scenario 1	Scenario 2	Pooled
Hygiene	-0.138 (0.084)	-0.083 (0.060)	-0.096** (0.045)
Structure	-0.001 (0.096)	-0.023 (0.049)	-0.005 (0.041)
CIM	-1.309 (64.981)	-0.086** (0.035)	-0.159*** (0.030)
Scenario 2 Dummy			-0.727* (0.384)
Cut Constants	Yes (6 levels)	Yes (6 levels)	Yes (6 levels)
Observations	34	35	69

*Notes: Dependent variable is Food Hygiene Rating. Standard errors are shown in parentheses. Significance levels for coefficients: *= p < 0.1, ** = p < 0.05, *** = p < 0.01.*

Table A2: Ordered Probit regression model, determinants of FH Rating.

	Scenario 1	Scenario 2	Pooled
gender (1 if female)	-0.049 (0.441)	0.109 (0.403)	0.045 (0.283)
age	0.074** (0.034)	0.014 (0.029)	0.028 (0.020)
experience	-1.268*** (0.460)	-0.026 (0.309)	-0.375* (0.218)
conscientious	-0.314 (0.566)	0.533 (0.511)	0.110 (0.353)
Cut Constants	Yes (6 levels)	Yes (6 levels)	Yes (6 levels)
Observations	34	34	68

*Notes: Dependent variable is Food Hygiene Rating. Standard errors are shown in parentheses. Significance levels for coefficients: *= p < 0.1, ** = p < 0.05, *** = p < 0.01.*

Table A3: Ordered Probit regression model, determinants of FH rating components.

	Hygiene	Structure	CIM
gender (1 if female)	-0.350 (0.291)	-0.793** (0.312)	-0.116 (0.293)
age	-0.031 (0.021)	0.000 (0.022)	-0.017 (0.021)
experience	0.386* (0.218)	-0.239 (0.230)	0.338 (0.223)
conscientious	0.031 (0.359)	0.424 (0.384)	-0.026 (0.364)
Cut Constants	Yes (6 levels)	Yes (6 levels)	Yes (6 levels)
Observations	67	67	68

*Notes: Dependent variable is specified at the top of the model column. Data is pooled from both scenarios. Standard errors are shown in parentheses. Significance levels for coefficients: *= p < 0.1, ** = p < 0.05, *** = p < 0.01.*

Table A4: Ordered Probit regression model, determinants of FH rating and components

	FH Rating	Hygiene	Structure	CIM
group2	-6.214	-1.691	-0.582	4.265
	(5.919)	(6.101)	(5.933)	(5.953)
gender (1 if female)	1.947*	-1.699*	-0.359	-2.430**
	(1.002)	(1.030)	(0.875)	(1.184)
age	0.030	-0.132	0.041	-0.045
	(0.061)	(0.083)	(0.047)	(0.064)
experience	-2.778**	4.755**	-0.425	3.454**
	(1.299)	(1.912)	(0.719)	(1.500)
conscientious	0.285	-1.211	0.248	-0.593
	(1.101)	(1.245)	(1.173)	(1.206)
group2*gender	-3.140**	2.075*	-1.224	2.584*
	(1.307)	(1.254)	(1.150)	(1.359)
group2*age	0.133	0.124	-0.084	-0.057
	(0.086)	(0.100)	(0.075)	(0.084)
group2*experience	0.956	-4.935**	0.398	-2.431
	(1.395)	(1.987)	(0.893)	(1.518)
group2*conscientious	0.027	0.930	1.335	0.465
	(1.405)	(1.531)	(1.531)	(1.457)
Cut Constants	Yes (4 levels)	Yes (3 levels)	Yes (2 levels)	Yes (2 levels)
Observations	34	33	33	34

*Notes: Dependent variable is specified at the top of the model column. Data is scenario 1 only. Standard errors are shown in parentheses. Significance levels for coefficients: *= p < 0.1, ** = p < 0.05, *** = p < 0.01.*

Table A5: Ordered Probit regression model, determinants of FH rating and components

	FH Rating	Hygiene	Structure	CIM
group2	-4.037	5.384	6.073	7.729
	(5.842)	(5.898)	(5.586)	(7.851)
gender (1 if female)	1.339	-1.563*	-0.461	-1.206
	(0.818)	(0.920)	(0.815)	(1.304)
age	-0.129***	-0.017	0.023	0.270*
	(0.046)	(0.044)	(0.044)	(0.146)
experience	0.957	0.527	-1.236*	0.982
	(0.691)	(0.695)	(0.707)	(0.852)
conscientious	2.027*	-0.479	1.660	-3.012
	(1.199)	(1.176)	(1.188)	(1.894)
group2*gender	-1.133	-0.449	-0.641	1.185
	(1.045)	(1.134)	(1.030)	(1.463)
group2*age	0.355***	-0.171**	-0.067	-0.320**
	(0.088)	(0.079)	(0.066)	(0.156)
group2*experience	-2.023**	0.448	1.221	-1.279
	(0.870)	(0.852)	(0.823)	(0.962)
group2*conscientious	-2.177	0.777	-1.410	2.828
	(1.428)	(1.427)	(1.405)	(2.044)
Cut Constants	Yes (5 levels)	Yes (3 levels)	Yes (2 levels)	Yes (3 levels)
Observations	34	34	34	34

*Notes: Dependent variable is specified at the top of the model column. Data is scenario 1 only. Standard errors are shown in parentheses. Significance levels for coefficients: *= p < 0.1, ** = p < 0.05, *** = p < 0.01.*

Table A6: Ordered Probit regression model, scenario 1 FH rating and components.

	FH rating	Hygiene	Structure	CIM
confidence	-1.486	2.752	-0.540	0.469
	(3.456)	(3.613)	(3.614)	(3.390)
gender (1 if female)	-0.568	0.034	-0.848	-0.155
	(0.550)	(0.515)	(0.530)	(0.488)
age	0.125***	-0.017	-0.006	-0.064*
	(0.044)	(0.041)	(0.039)	(0.037)
experience	-1.818***	0.456	-0.402	1.050**
	(0.613)	(0.423)	(0.428)	(0.421)
conscientious	-0.242	1.495*	1.042	-0.138
	(0.764)	(0.821)	(0.797)	(0.715)
confidence*gender	2.440***	1.014	0.427	-1.489**
	(0.807)	(0.755)	(0.728)	(0.700)
confidence*age	-0.029	-0.006	0.054	-0.020
	(0.049)	(0.052)	(0.052)	(0.049)
confidence*experience	0.079	0.486	0.152	0.536
	(0.665)	(0.637)	(0.603)	(0.643)
confidence*conscient.	0.314	-1.465	-0.788	0.128
	(0.791)	(0.899)	(0.867)	(0.777)
Cut Constants	Yes (4 levels)	Yes (3 levels)	Yes (2 levels)	Yes (2 levels)
Observations	34	33	33	34

*Notes: Dependent variable is specified at the top of the model column. Data is scenario 1 only. Standard errors are shown in parentheses. Significance levels for coefficients: *= p < 0.1, ** = p < 0.05, *** = p < 0.01.*

Table A7: Ordered Probit regression model, scenario 2 FH rating and components.

	FH rating	Hygiene	Structure	CIM
confidence	5.525	1.061	8.470**	-2.716
	(3.438)	(3.546)	(3.947)	(3.304)
gender (1 if female)	-0.157	0.594	-0.479	-0.145
	(0.523)	(0.538)	(0.593)	(0.519)
age	0.104**	-0.009	-0.021	-0.070*
	(0.041)	(0.039)	(0.040)	(0.036)
experience	-1.275**	0.557	-0.353	0.807**
	(0.516)	(0.431)	(0.507)	(0.410)
conscientious	-0.195	0.164	2.129**	0.157
	(0.692)	(0.772)	(0.965)	(0.691)
confidence*gender	1.746**	-0.509	-2.454**	-1.733**
	(0.814)	(0.774)	(1.111)	(0.811)
confidence*age	-0.052	0.003	0.063	0.018
	(0.052)	(0.052)	(0.050)	(0.049)
confidence*experience	0.541	1.371*	-0.232	0.076
	(0.820)	(0.812)	(0.758)	(0.765)
confidence*conscient.	-1.444	-1.093	-2.582**	0.888
	(0.968)	(1.008)	(1.154)	(0.940)
Cut Constants	Yes (4 levels)	Yes (3 levels)	Yes (2 levels)	Yes (2 levels)
Observations	34	33	33	34

*Notes: Dependent variable is specified at the top of the model column. Data is scenario 2 only. Standard errors are shown in parentheses. Significance levels for coefficients: *= p < 0.1, ** = p < 0.05, *** = p < 0.01.*

Table A8a: Ordered Probit regression model, scenario 1 rating and rationale.

	(1)	(2)
Paperwork	0.037	
	(0.375)	
New business		-0.342
		(1.063)
Cut Constants	Yes (4 levels)	Yes (4 levels)
Observations	35	35

*Notes: Dependent variable is rating. Data is scenario 1 only. Standard errors are shown in parentheses. Significance levels for coefficients: *= p < 0.1, ** = p < 0.05, *** = p < 0.01.*

Table A8b: Ordered Probit regression model, scenario 1 rating and rationale.

	Positive about Knowledge	Negative about Knowledge
Knowledge	0.293	0.421
	(0.580)	(0.495)
Cut Constants	Yes (4 levels)	Yes (4 levels)
Observations	28	31

*Notes: Dependent variable is rating; Selection of independent variable is specified at the top of the model column. Data is scenario 1 only. Standard errors are shown in parentheses. Significance levels for coefficients: *= p < 0.1, ** = p < 0.05, *** = p < 0.01.*

Table A8c: Ordered Probit regression model, scenario 1 rating and rationale.

	Positive about Risks	Negative about Risks
Risks	1.586	-0.162
	(1.084)	(0.463)
Cut Constants	Yes (4 levels)	Yes (4 levels)
Observations	28	34

*Notes: Dependent variable is rating; Selection of independent variable is specified at the top of the model column. Data is scenario 1 only. Standard errors are shown in parentheses. Significance levels for coefficients: *= p < 0.1, ** = p < 0.05, *** = p < 0.01.*

Table A8d: Ordered Probit regression model, scenario 1 rating and rationale.

	Positive about Hygiene	Negative about Hygiene
Hygiene	-1.040	-0.567
	(0.708)	(0.544)
Cut Constants	Yes (4 levels)	Yes (4 levels)
Observations	30	31

*Notes: Dependent variable is rating; Selection of independent variable is specified at the top of the model column. Data is scenario 1 only. Standard errors are shown in parentheses. Significance levels for coefficients: *= p < 0.1, ** = p < 0.05, *** = p < 0.01.*

Table A9a: Ordered Probit regression model, scenario 2 rating and rationale.

	(1)	(2)
Paperwork	0.113	
	(0.385)	
New business		0.258
		(0.476)
Cut Constants	Yes (5 levels)	Yes (5 levels)
Observations	35	35

*Notes: Dependent variable is rating. Data is scenario 2 only. Standard errors are shown in parentheses. Significance levels for coefficients: *= p < 0.1, ** = p < 0.05, *** = p < 0.01.*

Table A9b: Ordered Probit regression model, scenario 2 rating and rationale.

	Positive about Knowledge	Negative about Knowledge
Knowledge	0.759*	0.459
	(0.454)	(0.483)
Cut Constants	Yes (5 levels)	Yes (5 levels)
Observations	28	31

*Notes: Dependent variable is rating; Selection of independent variable is specified at the top of the model column. Data is scenario 2 only. Standard errors are shown in parentheses. Significance levels for coefficients: *= p < 0.1, ** = p < 0.05, *** = p < 0.01.*

Table A9c: Ordered Probit regression model, scenario 2 rating and rationale.

	Positive about Risks	Negative about Risks
Risks	0.021	0.125
	(0.499)	(0.439)
Cut Constants	Yes (5 levels)	Yes (5 levels)
Observations	28	34

*Notes: Dependent variable is rating; Selection of independent variable is specified at the top of the model column. Data is scenario 2 only. Standard errors are shown in parentheses. Significance levels for coefficients: *= p < 0.1, ** = p < 0.05, *** = p < 0.01.*

Table A9d: Ordered Probit regression model, scenario 2 rating and rationale.

	Positive about Hygiene	Negative about Hygiene
Hygiene	0.435	0.324
	(0.380)	(0.361)
Cut Constants	Yes (5 levels)	Yes (5 levels)
Constant	0.733***	0.762***
	(0.280)	(0.285)
Observations	30	31

*Notes: Dependent variable is rating; Selection of independent variable is specified at the top of the model column. Data is scenario 2 only. Standard errors are shown in parentheses. Significance levels for coefficients: *= p < 0.1, ** = p < 0.05, *** = p < 0.01.*

Table A10a: Ordered Probit regression model, scenario 1 rating and follow-up.

	(1)	(2)
Re-visit	-0.646	
	(0.424)	
Provide Coaching		-0.142
		(0.384)
Cut Constants	Yes (4 levels)	Yes (4 levels)
Observations	35	35

*Notes: Dependent variable is rating. Data is scenario 1 only. Standard errors are shown in parentheses. Significance levels for coefficients: *= p < 0.1, ** = p < 0.05, *** = p < 0.01.*

Table A10b: Ordered Probit regression model, scenario 1 CIM score and follow-up.

	(1)	(2)
Re-visit	0.717	
	(0.441)	
Provide Coaching		-0.078
		(0.394)
Cut Constants	Yes (4 levels)	Yes (4 levels)
Observations	35	35

*Notes: Dependent variable is CIM score. Data is scenario 1 only. Standard errors are shown in parentheses. Significance levels for coefficients: *= p < 0.1, ** = p < 0.05, *** = p < 0.01.*

Table A11a: Ordered Probit regression model, scenario 2 rating and follow-up.

	(1)	(2)
Re-visit	-0.507	
	(0.372)	
Provide Coaching		-0.244
		(0.371)
Cut Constants	Yes (5 levels)	Yes (5 levels)
Observations	35	35

*Notes: Dependent variable is rating. Data is scenario 2 only. Standard errors are shown in parentheses. Significance levels for coefficients: *= p < 0.1, ** = p < 0.05, *** = p < 0.01.*

Table A11b: Ordered Probit regression model, scenario 2 CIM score and follow-up.

	(1)	(2)
Re-visit	1.197**	
	(0.491)	
Provide Coaching		-0.096
		(0.410)
Cut Constants	Yes (3 levels)	Yes (3 levels)
Observations	35	35

*Notes: Dependent variable is CIM score. Data is scenario 2 only. Standard errors are shown in parentheses. Significance levels for coefficients: *= p < 0.1, ** = p < 0.05, *** = p < 0.01.*

Table A12: Ordered Probit regression model, scenario 1 colleagues rating.

	(1)
Confidence in own score	0.707***
	(0.259)
Cut Constants	Yes (3 levels)
Observations	34

*Notes: Dependent variable is estimated percentage of colleagues that would have given the same rating; independent variable is confidence in own rating. Data is scenario 1 only. Standard errors are shown in parentheses. Significance levels for coefficients: *= p < 0.1, ** = p < 0.05, *** = p < 0.01.*

Table A13: Ordered Probit regression model, scenario 2 colleagues rating.

	(1)
Confidence in own score	0.898***
	(0.291)
Cut Constants	Yes (3 levels)
Observations	35

*Notes: Dependent variable is estimated percentage of colleagues that would have given the same rating; independent variable is confidence in own rating. Data is scenario 2 only. Standard errors are shown in parentheses. Significance levels for coefficients: *= p < 0.1, ** = p < 0.05, *** = p < 0.01.*

Appendix B: Materials

Scenario 1: Inspection simulation

In this part of the study, you will be carrying out an inspection of a kitchen in the Saw Swee Hock building on the LSE campus. Please carry out the inspection as you usually would, but with the following changes of procedure:

1. The inspection will take **30 minutes** maximum. We appreciate that a typical inspection will usually take longer than that, but due to time constraints you will be limited to 30 minutes only.
2. You will **not deliver your final verdict in person** at the end of the inspection. Instead, we ask you to complete the documentation for the inspection afterwards, before you start the next part of the study. We appreciate that providing feedback is an important part of the inspection, but due to time constraints this will not be possible. We will provide you with an opportunity to submit your feedback in writing afterwards.

We will provide you with a notepad and report sheet for the inspection. Please complete the report sheet after the inspection, as well as the feedback form that accompanies it.

Please write notes and score the establishment **as you would if you were conducting a real inspection** – this is very important for the scientific validity of today's study.

Please read the 'Case file' on the next page before you start the inspection.

Scenario 1: Case file

You will be inspecting a kitchen of a small contract caterer that prepares food that is sold in the Student Union canteen in the same building. You have 30 minutes for the inspection, at which point you should conclude the inspection and make your way to an adjoining room for part 2 of the study. The staff in the kitchen or one of our research colleagues will remind you when 30 minutes have passed.

The kitchen is staffed by a kitchen assistant and a chef. The chef is responsible for the running of the kitchen and will be your point of contact during the inspection. This is the current chef's first week in the kitchen. Note that he chef is only in charge of one half of the kitchen space – you can ignore the other half for the purposes of the inspection.

Please assume that you have already spoken to the manager and that he/she will not be available during the inspection of the kitchen. This means you will have to base your assessment of the kitchen on your conversations with the chef. The manager has shown you training records for the kitchen staff and historical food safety records for the past 6 months. The chef keeps the records for the current month in the kitchen.

Historical record (group 1):

You know that this establishment was inspected 6 months ago. There were significant hygiene/structure/management non-compliances during this inspection, resulting in subsequent visits. **At the last intervention, the establishment received a Food Hygiene Rating of 2.**

Historical record (group 2):

You know that this establishment was inspected 18 months ago. There were no significant non-compliances found during this previous inspection. **During the previous inspection, the establishment received a Food Hygiene Rating of 5.**

Scenario 2: Inspection simulation

In this part of the study, you will be carrying out an inspection of another kitchen in the Saw Swee Hock building on the LSE campus. Please carry out the inspection as you usually would, but with the following changes of procedure:

1. The inspection will take **30 minutes** maximum. We appreciate that a typical inspection will usually take longer than that, but due to time constraints you will be limited to 30 minutes only.
2. You will **not deliver your final verdict in person** at the end of the inspection. Instead, we ask you to complete the documentation for the inspection afterwards, before you start the next part of the study. We appreciate that providing feedback is an important part of the inspection, but due to time constraints this will not be possible. We will provide you with an opportunity to submit your feedback in writing afterwards.

We will provide you with a notepad and report sheet for the inspection. Please complete the report sheet after the inspection, as well as the feedback form that accompanies it.

Please write notes and score the establishment **as you would if you were conducting a real inspection** – this is very important for the scientific validity of today's study.

Please read the 'Case file' on the next page before you start the inspection.

Scenario 2: Case file

You will be inspecting a kitchen of a small contract caterer that prepares food for functions and receptions on the LSE campus. You have 30 minutes for the inspection, at which point you should conclude the inspection and make your way to an adjoining room for part 4 of the study The staff in the kitchen or one of our research colleagues will remind you when 30 minutes have passed.

The kitchen is staffed by a kitchen assistant and a chef. The chef is responsible for the running of the kitchen and will be your point of contact during the inspection. This is the current chef's first week in the kitchen. Note that he chef is only in charge of one half of the kitchen space – you can ignore the other half for the purposes of the inspection.

Please assume that you have already spoken to the manager and that he/she will not be available during the inspection of the kitchen. This means you will have to base your assessment of the kitchen on your conversations with the chef. The manager has shown you training records for the kitchen staff and historical food safety records for the past 6 months. The chef keeps the records for the current month in the kitchen.

Historical record (both groups):

As this is the first time that this establishment is inspected, **you have no previous records for it**.

Scenario 1 and 2: post-scenario feedback form

Thank you for completing the inspection. Please ensure that you have completed the entire report sheet. To finish the inspection, could you please answer the questions below. When you have finished answering these questions, please put this form and report sheet in the white envelope provided and place it in the 'Inspection forms' tray at the entry to the Behavioural Research Lab.

1. If you had had more time for the inspection, how would you have spent it?

2. If you had had time to sit down with the chef and provide feedback in person, what would your general feedback have been?

3. If you had had time to sit down with the chef and provide feedback in person, which issues would you have raised?

4. If you were employed by the Local Authority for this establishment, what other actions would you have taken after this inspection?

Scenario 1: post-scenario questionnaire

1. Confidence scale

	1- Not at all confident	2	3	4	5 - Extremely confident
Overall, how confident are you in the rating of the premises just examined?					
Overall, how confident are you in your abilities as a food safety EHO					

2. Cues influencing decision-making

Please report below the three observations that most influenced your rating for the premises just examined.
1.
2.
3.

What was the single most important factor that influenced your final assessment

3. Colleagues

Item	0-20%	20-40%	40-60%	60-80%	80-100%
What percentage of your colleagues (other EHOs) do you think would have given the same rating as you?					

Why might some EHOs have made a different assessment to you?

4. Affective Scale

This scale consists of a number of words and phrases that describe different feelings and emotions. . Indicate to what extent you feel this way at the moment. Read each item and then circle the appropriate answer on the 5-point scale.

Item	Very slightly	A little	Moderately	Quite a bit	Extremely
Interested	1	2	3	4	5
Distressed	1	2	3	4	5
Excited	1	2	3	4	5
Upset	1	2	3	4	5
Strong	1	2	3	4	5
Guilty	1	2	3	4	5
Scared	1	2	3	4	5
Hostile	1	2	3	4	5
Enthusiastic	1	2	3	4	5
Proud	1	2	3	4	5
Irritable	1	2	3	4	5
Alert	1	2	3	4	5
Ashamed	1	2	3	4	5
Inspired	1	2	3	4	5
Nervous	1	2	3	4	5
Determined	1	2	3	4	5
Attentive	1	2	3	4	5
Jittery	1	2	3	4	5
Active	1	2	3	4	5
Afraid	1	2	3	4	5

5. Demographics

	0-1 years	1-2 years	2-5 years	5-10 years	10+ years
How many years' experience do you have as an EHO?					

	0-6 months	6-12 months	12-24 months	2-5 years	5+ years
When was the last time you went on a training course?					

Please list your qualifications relevant to your job.

	Male	Female
Please indicate your gender		

Please indicate your age	

	Yes	No
Have you inspected a university facility before?		

6. 90% Confidence interval questions. For each of the following questions, please answer them by specifying the minimum and maximum points of a 90% confidence interval. This means that you should specify two values such that you are 90% certain that the true answer lies between them. For example, for the question "In what year was Winston Churchill born?" you should specify two dates (e.g. 1870 and 1900) such that you are 90% certain that the true answer is between those dates.

	Minimum	Maximum
In what year did the 'horsemeat scandal' happen?		

	Minimum	Maximum
In what year did the 'Mad Cow disease' (BSE) epidemic start in the UK?		

	Minimum	Maximum
What is the recommended minimum temperature (in degrees Celsius) for cooking chicken?		

	Minimum	Maximum
What percentage of food poisoning cases occur in the home?		

	Minimum	Maximum
What percentage of chicken meat sold in stores contains the salmonella bacteria?		

Scenario 2: post-scenario questionnaire

1. Confidence scale

	1- Not at all confident	2	3	4	5 - Extremely confident
Overall, how confident are you in the rating of the premises just examined?					
Overall, how confident are you in your abilities as a food safety EHO					

2. Cues influencing decision-making

Please report below the three observations that most influenced your rating for the premises just examined.
1.
2.
3.

What was the single most important factor that influenced your final assessment.

3. Colleagues

Item	0-20%	20-40%	40-60%	60-80%	80-100%
What percentage of your colleagues (other EHOs) do you think would have given the same rating as you?					

Why might some EHOs have made a different assessment to you?

4. Industry

Item	1*	2*	3*	4*	5*
Across the entire UK, indicate the percentage of catering facilities that receive each rating score (total should add up to 100%)					

5. Affective Scale.

This scale consists of a number of words and phrases that describe different feelings and emotions. Indicate to what extent you feel this way at the moment. Read each item and then circle the appropriate answer on the 5-point scale.

Item	Very slightly	A little	Moderately	Quite a bit	Extremely
Interested	1	2	3	4	5
Distressed	1	2	3	4	5
Excited	1	2	3	4	5
Upset	1	2	3	4	5
Strong	1	2	3	4	5
Guilty	1	2	3	4	5
Scared	1	2	3	4	5
Hostile	1	2	3	4	5
Enthusiastic	1	2	3	4	5
Proud	1	2	3	4	5
Irritable	1	2	3	4	5
Alert	1	2	3	4	5
Ashamed	1	2	3	4	5
Inspired	1	2	3	4	5
Nervous	1	2	3	4	5
Determined	1	2	3	4	5
Attentive	1	2	3	4	5
Jittery	1	2	3	4	5
Active	1	2	3	4	5
Afraid	1	2	3	4	5

6. Statements.

This scale consists of a number of statements about yourself. Read each item and then circle the appropriate answer on the 5-point scale.					
Item	Strongly disagree	Disagree	Neither	Agree	Strongly Agree
I am always prepared	1	2	3	4	5
I pay attention to details	1	2	3	4	5
I waste my time	1	2	3	4	5
I get chores done right away	1	2	3	4	5
I carry out my plans	1	2	3	4	5
I make plans and stick to them	1	2	3	4	5
I find it difficult to get down to work	1	2	3	4	5
I do just enough work to get by	1	2	3	4	5
I don't see things through	1	2	3	4	5
I complete tasks successfully	1	2	3	4	5
I shirk my duties	1	2	3	4	5
I mess things up	1	2	3	4	5
I do things according to a plan	1	2	3	4	5
I am exacting in my work	1	2	3	4	5
I leave things unfinished	1	2	3	4	5
I don't put my mind on task at hand	1	2	3	4	5
I finish what I start	1	2	3	4	5
I follow through with my plans	1	2	3	4	5
I make a mess of things	1	2	3	4	5
I need a push to get started	1	2	3	4	5

7. Open questions.

In your opinion, what are the most important skills for an EHO?

Overall, what is the single most important factor that influences your assessment of a catering premises?

Have you any other comment (e.g. on the scenarios, factors that influence the decision-making of EHOs)?

Scenario 1 and 2: inspection form

FOOD BUSINESS ESTABLISHMENT REPORT

1. Name of EHO:

2. Trading name of business:

3. Name of person seen/interviewed:

4. Date and time of inspection:

5. Areas inspected/audited:

6. Documents and other records examined:

7. Key points discussed during the visit:

8. Non-compliances observed:		
Hygiene	Structure	Confidence in management

9. Actions to be taken by the FBO :

10. Assessment scores:		
Hygiene (0-25)	Structure (0-25)	Confidence in management (0-30)

11. Final Food Hygiene Rating assessment (0-5 stars):

11b. Final Food Hygiene Rating rationale (optional).

Pre-study: Informed consent form

Informed Consent Form

Today, we will ask you to participate in a 4-part study. In two parts of the study (part 1 and 3) you will be conducting a simulation of a kitchen inspection; the other two parts (part 2 and 4) are questionnaire based.

During our study, we will be collecting data from you by asking you to complete various forms and questionnaires – this form is to ask for your consent for this data to be collected. Once you have read the entire consent form, and if you are willing to participate, please sign at the bottom to indicate that you have understood everything and agree to take part.

Please read this consent form carefully before you decide whether or not you want to participate in this research study. You are free to ask questions at any time before, during, or after your participation. Should you at any time choose to withdraw your consent, you will be allowed to do so, and all evidence of your participation will be promptly destroyed.

We will be collecting data on demographic information, your attitudes and approach to work, as well as your responses in the two simulated scenarios. Please note that all of this data will serve research purposes only and is in no way intended to be a measure of individual performance.

All of your answers will remain completely confidential and anonymous. Any storage, analysis or archival of the data by us (the researchers) will exclude your name or other personal information. Your personal details will only be used by us to report your participation to the FSA and to reimburse you for your expenses.

There are no risks to you from this research and no foreseeable direct benefits. It is hoped that the research will benefit policy makers and the design of consistency training for EHOs.

At the end of today, you can let us know if you would like to be informed about the purpose and outcome of this study. Once the study has been concluded, we will share the results with you.

****

I have read and understand this consent form and am willing to take part in this research study.

_____ _____

Signature Name (please print)

APPENDIX C: PHOTOS.

Scenario 1

Scenario 2