CEBRA Project 1404C:
Testing Compliance-Based Inspection Protocols

Supplementary Report

Anthony Rossiter[1,2], Andreas Leibbrandt[3] and Susie Hester[4,5]

1. Centre for Market Design, Victorian Department of Treasury and Finance and The University of Melbourne
2. Department of Econometrics and Business Statistics, Monash University
3. Department of Economics and Monash Experimental Research Insights Team, Monash University
4. CEBRA, The University of Melbourne
5. UNE Business School, University of New England

7 July 2018

MONASH
BUSINESS
SCHOOL

Acknowledgements

This report is a product of the Centre of Excellence for Biosecurity Risk Analysis (CEBRA). In preparing this report, the authors acknowledge the financial and other forms of support provided by the Department of Agriculture and Water Resources, the University of Melbourne, the Victorian Department of Treasury and Finance, Monash University and the University of New England.

The authors are grateful to the following people who generously gave their time to attend the project workshop in November 2015: Lois Ransom, Anthony Wicks, Jenni Edwards, Greg Hood, David Mackay, Andrew Trainer and Bo Wang (Commonwealth Government Department of Agriculture and Water Resources); Tony Arthur and Phil Tennant (Australian Bureau of Agricultural and Resource Economics and Sciences, Commonwealth Government Department of Agriculture and Water Resources); and Mark Burgman (CEBRA).

We thank Joseph Vecci and Kevin Wu (Monash University) for their assistance in running the experimental sessions, advice on the experiment format and programming the computer activities for the experiments using the Zurich Toolbox for Readymade Economics Experiments (z-Tree). The final report and experiments benefited significantly from discussions at various stages with Charlie Plott (California Institute of Technology); John List (University of Chicago); Mark Burgman, Andrew Robinson and Jessica Holliday (CEBRA); Lana Friesen (The University of Queensland); Udeni Perera (Monash University); and Gary Stoneham, Blair Cleave and Reiss McLeod (Centre for Market Design, Victorian Department of Treasury and Finance). Comments from participants at the inaugural Environmental Economics Workshop, held at Monash University, and the Tenth Annual Australia New Zealand Workshop on Experimental Economics, held at the University of Tasmania, on preliminary versions of this work are also gratefully acknowledged. The authors also appreciate the comments provided by two anonymous scientific reviewers and two anonymous reviewers from the Department of Agriculture and Water Resources, which have helped improve this report.

Finally, we acknowledge the 275 participants in our experimental sessions at the Monash University Laboratory for Experimental Economics, without whom this work would not be possible.

Table of Contents

Table of Definitions

Approach rate: An estimate of the likelihood of entry of pests and diseases determined through inspection results.

Biosecurity risk material: Material that has the potential to introduce a pest or disease to Australia. This could include, but is not limited to: live insects, seeds, soil, dirt, clay, animal material, and plant material such as straw, twigs, leaves, roots, bark, food refuse and other debris.

Clearance number: A key parameter of the CSP-1 and CSP-3 algorithms. It represents the number of consecutive clean lines that must be reached before a target's goods can be switched to a reduced inspection rate (i.e. switched to monitoring mode).

Confidence interval: A type of interval estimate, computed from observed (sample) data, that might contain the true value of an unknown population parameter. This interval estimate has an associated *confidence level*, which represents the proportion of possible confidence intervals that would contain the true value of the unknown population parameter. This interval estimate assumes, hypothetically, that an infinite number of independent samples of the same size as the observed data could be drawn, allowing relevant sample statistics to be calculated. Most confidence intervals presented in this report use a 95 per cent confidence level.

Consignment: In general, a consignment consists of all the goods for a single consignee that arrives on the same voyage of a vessel; a single consignment can consist of many container loads of goods.

CSP (continuous sampling plan): A technical rule for determining whether or not to inspect a consignment, based on the recent inspection history of the pathway and some parameters the pathway manager sets. (Dodge and Torrey, 1951).

Economics experiment: An economics experiment can refer to several related research methods used to collect data for scientific purposes so as to understand the factors that influence people's decisions in economically relevant situations, either as individuals or in a group setting. A key commonality of these approaches is that the researcher maintains some control over the environment of interest and/or the allocation of participants to treatments (see below). A *conventional laboratory experiment* is conducted in a computer laboratory with university students, while a *field experiment* is characterised by augmenting the laboratory experiment with elements from the natural context for studying interactions with rules and institutions.

(Experimental) treatment: Each treatment represents a specific combination of the collection of characteristics analysed in the experiment. In this experiment, the characteristics include: the type of inspection rule (CSP-1 or CSP-3); the clearance number and monitoring fraction of the inspection rule; the level of information provided to participants about the rule; the nature of feedback given to participants; the costs incurred in being inspected or treated; and whether the participant has a choice over the rule they follow. The results from different treatments can be compared only where one of these characteristics is varied at a time, with all others held constant.

Framing: Relates to the presentation of information that shifts the perspective of decision-makers in ways that can change the way they evaluate alternative options. (Weber, 2013, 387).

Heuristic (technique): A mental shortcut applied in problem solving, learning or discovery to help arrive at a decision in a context where finding the optimal solution is challenging, impractical or impossible. Practical methods drawing on selected salient features of the problem are usually employed, though these are not guaranteed to be optimal or perfect.

Implied approach rate: An estimate of the approach rate for consignments in the main experimental task. This is a weighted average of the biosecurity risk material approach rates of the available suppliers, weighted by the number of choices of that supplier made by participants in each treatment.

$$\textit{Implied approach rate (\%)} = \frac{\sum_{i=A,B,C,D} \text{Choices for supplier i} \times \text{Approach rate for supplier i (\%)}}{\text{Total supplier choices}}$$

The number of choices in the above formula can be either taken at a particular time point (period) or aggregated across periods in the multi-period task.

Inspection: Examination of product or systems for the biosecurity of animal, plant, food and human health to verify that they conform to requirements (Beale, 2008).

Inspection failure: In general, an inspection failure occurs when there is a non-compliance detected at inspection. The possible types of non-compliance include the incorrect declaration of goods, packaging failures and the presence of biosecurity risk material in consignments. For the purposes of the experiment, it is assumed all inspection failures are due to the presence of biosecurity risk material in consignments.

Inspection game: A mathematical model of a situation where an *inspector* verifies that another party (the *inspectee*) adheres to certain legal requirements (Avenhaus et al., 2002, 1949).

Institution: The set of rules or procedures that govern how different agents can interact in an economic system.

Intervention: Legally enforceable obligations (through legislation or regulations) imposed by government on business and/or the community, together with government administrative processes that support the obligations. In the biosecurity context, this includes requirements related to:

- prescribing specific actions that must be completed before goods can be brought into Australia;
- giving notice of goods to be unloaded in Australian territory;
- providing information, including documents, about the goods if requested by biosecurity officers;
- allowing for the goods to be physically inspected;
- allowing for samples of the goods to be taken; and
- prescribing treatments for rectifying the presence of biosecurity risk material in a consignment.

Monitoring fraction: A parameter in the CSP-1 and CSP-3 rule used to determine the frequency of inspection once an importer has demonstrated sufficient compliance with biosecurity requirements in the monitoring mode of the CSP algorithm. This parameter governs the reduced rate of inspection (MF) to be applied that enables inspection of less than 100% of consignments imported.

Natural framing: Refers to the experimental instructions (or script) being prepared in a way that describes the real-world context underpinning the experimental study. The opposite of abstract framing, where the instructions are devoid of the real-world context for the experiment.

Period: The unit of time for a sequence of repeated decision processes in an experiment. In multi-period tasks, experimental subjects make choices based on the same set of rules and/or parameters as part of the replication process.

Power (statistical power): For a binary test of hypotheses, the power is the probability that the test correctly rejects the null hypothesis (H_0) when the alternative hypothesis (H_1) is true. Where the null hypothesis is that there is no treatment effect, or a covariate has no effect on the choices made by experimental subjects, the power represents the ability of a statistical test to detect an effect if the effect actually exists.

P-value (probability value): For a binary test of hypotheses based on a given statistical model, the p-value is the probability that the statistical summary measure would be the same as or of greater magnitude than the actual observed results assuming the null hypothesis (H_0) is true.

For the purposes of this report, we mainly refer to two-sided hypothesis tests to see whether a treatment effect is different from zero. In this case, the p-value assesses the probability of a test statistic (usually a Z-score in most contexts for this report) being obtained that is the same as or larger in *absolute value* than the one computed for the given statistical model.

Tight census: A parameter in the CSP-3 algorithm which governs the number of consignments inspected at a rate of 100% following an inspection failure when the importer is in monitoring mode.

Treatment: Refers to actions required to rectify consignments found to contain biosecurity risk material during an inspection so they can be brought into Australia.

Treatment cost: The costs incurred by an importer resulting from treatments required by the biosecurity regulator to address the presence of biosecurity risk material in a consignment and allow the consignment to enter Australia.

1. Introduction and Overview of Report Structure

This report provides additional material to supplement the *Final Report* for *CEBRA Project 1404C: Testing Compliance-Based Protocols*. CEBRA Project 1404C sought to test the appropriateness of particular aspects of candidate mechanisms in a controlled environment using human subjects. These experiments sought to mimic the interactions between the department and importers relating to biosecurity inspections and employed well-established methodologies in experimental economics to ensure scientific validity around their findings.

The overall aim of this and related projects is to enhance the risk-return methodology of assessment currently being rolled out by the department. Refinements proposed in the sequences of projects (CEBRA Projects 1304C, 1404C and 1608C) consider the impact that different incentives and interventions have on the behaviour of import-supply chain participants and the resulting biosecurity risk of their consignments. Through the laboratory experiments in this project, the aim was to inform the department about critical aspects associated with implementing compliance-based protocols and identify how current practices may be fine-tuned to better support departmental objectives. It also serves as preparation for the field trial (CEBRA Project 1608C) by offering the opportunity to refine candidate mechanisms in a safe, low-cost environment before they are implemented in the field.

This experimental investigation drew on aspects of microeconomic theory discussed in the report on *CEBRA Project 1304C: Incentives for Importer Choices* (Rossiter et al., 2016). It also explored ways to improve the performance and efficacy of incentive-based frameworks by using ideas from behavioural economics to account for observed patterns of human behaviour in decision environments.

This supplementary report provides further context related to the design and theoretical foundations of the laboratory experiments and helps establish the scientific rigour around how the experimental data is analysed. The remainder of the report is structured as follows. Where relevant, cross-reference this material to the relevant sections of the *Final Report*.

Chapter 2 discusses experimental economics research methods in more detail than in Chapter 2.2 of the *Final Report*. It considers some key features associated with implementing experiments in economics and how they were applied in this set of experiments. This chapter also discusses the role of economics experiments in the economic design methodology and examines the features that differentiate laboratory from field experiments. In closing, it considers how results from laboratory experiments can be used to inform the specific policy context of interest.

In **Chapter 3** we document the theoretical underpinnings of the experimental framework in more detail than provided in Chapter 3.2 of the *Final Report*. It describes the types of factors influencing the department's choices and the behaviour of their stakeholders, drawing heavily on a modified version of the inspection game model of Rossiter and Hester (2017). We also consider the impact of using an individual-choice experiment to assess behaviour in response to different regulatory environments and outline some potential alternatives for assessing other aspects of importer behaviour.

The methods used to arrive at theoretical predictions of behaviour through simulation models are also discussed in detail, as is the approach adopted to calibrate specific parameters for various experimental treatments. The chapter also considers the role

that individual risk preferences may play in the supplier choices made by subjects in the experiment and describes the way in which risk preferences can be elicited in an economics experiment.

Chapter 4 is the first chapter discussing the experimental results in detail. This chapter previews aspects of the experimental data across treatments. A particular focus of this chapter is on describing demographic characteristics of the experimental subjects, the choices they made in the risk-elicitation task and their responses to the post-experiment questionnaire.

The main focus of **Chapter 5** is to chronicle the experimental results relating to treatment effects, as discussed in Chapter 4.1 to 4.6 inclusive, and check the robustness of findings presented in those sections of the *Final Report*. We use panel data regression models to assess whether the observed differences between treatments reflect actual treatment differences or can be explained by differences in subject characteristics. This section goes through the statistical methodology used to assess the results in this report before considering each pairwise comparison of interest in detail.

Chapter 6 considers the influence of individual characteristics on the supplier choices made by experimental subjects, the results of which were summarised in Chapter 4.7 of the *Final Report*. This analysis is underpinned by the statistical framework described in Chapter 5 of this report.

To complement the main matter of this *Supplementary Report*, there are two technical appendices as follows.

Appendix A documents the experimental instructions included in the computer-based instructions given to participants in z-Tree (Fischbacher, 2007), as well as instructions for the paper-based task at the end of the experiment.

Appendix B provides the MATLAB files used to simulate the payoff functions for importers in the individual-choice experiment setting. This relates to the calibration discussion included in Chapter 3.5 of this report.

2. Economics Experiments and Their Use in Public Policy

As noted by Smith (1994), Roth (1995) and Hertwig and Ortmann (2001), economics experiments have been used for a variety of different purposes, including:

- testing the predictions of formal theories, usually game-theoretic or decision-theoretic models, in a controlled environment to allow for unambiguous interpretation of the results;
- exploring the cause, or causes, of a theory's failure;
- isolating the cause, or causes, of observed experimental regularities, particularly for variables where there is little guidance from established theories;
- investigating the robustness of specific institutions by comparing different environments using the same institution; and
- exploring the effect of changes to the design of a particular market or institution through constructing an experimental environment that closely resembles key features of the naturally occurring environment of interest.

While the focus on this report is an experiment that seeks to deal largely with the last of these listed purposes of creating or fine-tuning institutions, doing so successfully means drawing on attributes that seek to address other roles of economics experiments.

This chapter reviews the use of economics experiments in public policy, such as in designing auctions, markets, contracts or regulation. It first outlines the basic setup of economics experiments, together with their characteristics and procedures,[1] before considering how these can be used to test the design of various institutions in public policy. In closing, we consider how the results of experiments in the laboratory and field can be used to inform decisions in the policy context.

Because our experiment is performed in the context of designing regulatory frameworks, we also consider how insights from behavioural economics can be used to improve the performance of institutions and how experiments can be used to trial some of these innovations.

2.1 Laboratory experiments

2.1.1 Recreating economic systems in the laboratory

The techniques applied in experimental economics allow researchers to examine self-contained microeconomic systems in a way that enables the behaviour of agents to be observed over time. These economies consistent of agents, each with their own characteristics, and institutions, or sets of rules, that govern how these agents can interact in the system.

Most experimental interactions in economics occur in an artificially created system in a computer laboratory, as was done in CEBRA Project 1404C. In this situation, the

[1] See Chapter 1 of Guala (2005) and Chapter 6 of Friedman and Cassar (2004) for more detailed explanations of some practical considerations in running economics experiments in a laboratory setting.

experimenter attains a very high degree of control over the experimental environment and enlists volunteers to make decisions in this controlled environment. In particular, Friedman and Cassar (2004, 25-26) note that, for each experimental subject, the experimenter can specify:

- the type of agent they are, such as whether they are an importer, supplier or regulator in the biosecurity system;
- the resources available to each agent, in terms of goods or time, at the start of the interactions (that is, their endowment);
- the possible choices each agent can make in the interactions, such as the suite of suppliers from which an importer could purchase particular goods;
- agents' preferences over different outcomes from interactions within the economic environment, represented by an outcome or payoff function; and
- the information agents know about others' endowments, preferences or the outcomes of interactions.

The controlled variation possible in this environment facilitates tests of theory and causality and identification of treatment effects (Falk and Heckman, 2009).

2.1.2 The role of economic theory and institutions in experiments[2]

An experimenter's selection of the factors they specify and control overlay a strong theoretical structure for economic experiments, enabling predictions to be made about what agents in the self-contained economy will do. The unifying framework of optimising behaviour, self-interest, rationality and notions of equilibrium that together underpin neoclassical economics allows the experimenter to predict each agent's actions and the expected overall outcome of the experiment. These predictions can then be tested experimentally and differences in outcomes from changing specific characteristics under the experimenter's control can also be compared. In the context of the biosecurity regulation experiments conducted in this project, Chapter 3 of this report outlines the theoretical underpinnings of these experiments.

Furthermore, strong institutions, such as markets, place significant constraints on the behaviour of agents. In contrast, individual-choice tasks, which is the domain of these experiments, provide weaker constraints, implying personal preferences can play a more prominent role in determining behaviour. The effect of institutions can still be examined under these circumstances, though additional information on underlying personal attributes and preferences should be collected from subjects to account for influences these may have on decision-making and performance in the experiment.

The important roles that theory and institutions play in economics experiments are two factors that are sometimes attributed to the differences in laboratory methods between experimental economics and experimental psychology (Friedman and Cassar, 2004, 18). Table 1 highlights further key differences in methodology salient for the economics experiments documented in this report.

[2] For a more in-depth discussion of contemporary views in the literature on this topic, see the book chapters in Part II of the collected volume edited by Frechette and Schotter (2015).

Table 1: Differences between typical economics and cognitive/social psychology experiments

Characteristic	Economics experiments	Psychology experiments
Instructions	Precisely defined (usually written) instructions ("scripts") are provided. These describe a subject's role, their action choices and the possible payoffs.	Scripts are not normally provided. Participants often do not assume clearly specified roles and are forced to "ad-lib", inferring what choices are available in the given situation.
Task repetition	Involve repeated tasks so participants can learn about the experimental environment and task, including how their choices and those of others may interact.	Behavioural decision-making studies tend to focus on "snapshot" studies that provide little or no opportunity for learning.
Financial incentives	Subjects are paid in cash based on clearly defined performance criteria that relate to the decisions they make in the experimental task.	Financial incentives are seldom used; when they are, they tend to be flat fees unrelated to performance in the experiment.
Deception of participants	Deception is considered taboo, with it being used in very few studies.	Deception is prevalent in social psychology experiments, in part to avoid "strategic behaviour" or investigate situations that may not occur naturally.

Source: Based on Hertwig and Ortmann (2001) and Friedman and Cassar (2004).

2.1.3 Instructions in economics experiments

As noted in Table 1, the vast majority of economics experiments follow a specified script to ensure consistency between experimental sessions and replicability of the experiments. The biosecurity experiment sessions used a mix of verbal instructions given by one of the instructors and written instructions delivered via the computer[3] or handouts in private to each participant in the experiment. Appendix A provides the computer- and paper-based instructions used in the experiment,[4] while Figures 1 and 2 provide screenshots of what experimental participants would see on their computer screens for the experimental instructions and when they make choices in z-Tree.

[3] The computer based activities forming part of the experiments were programmed and delivered using the Zurich Toolbox for Readymade Economic Experiments (z-Tree) – a widely used software package for developing and carrying out economic experiments (Fischbacher, 2007).

[4] The practice of providing experimental instructions and sharing experimental data is commonplace in experimental economics to allow for the results of particular experiments to be replicated by other researchers.

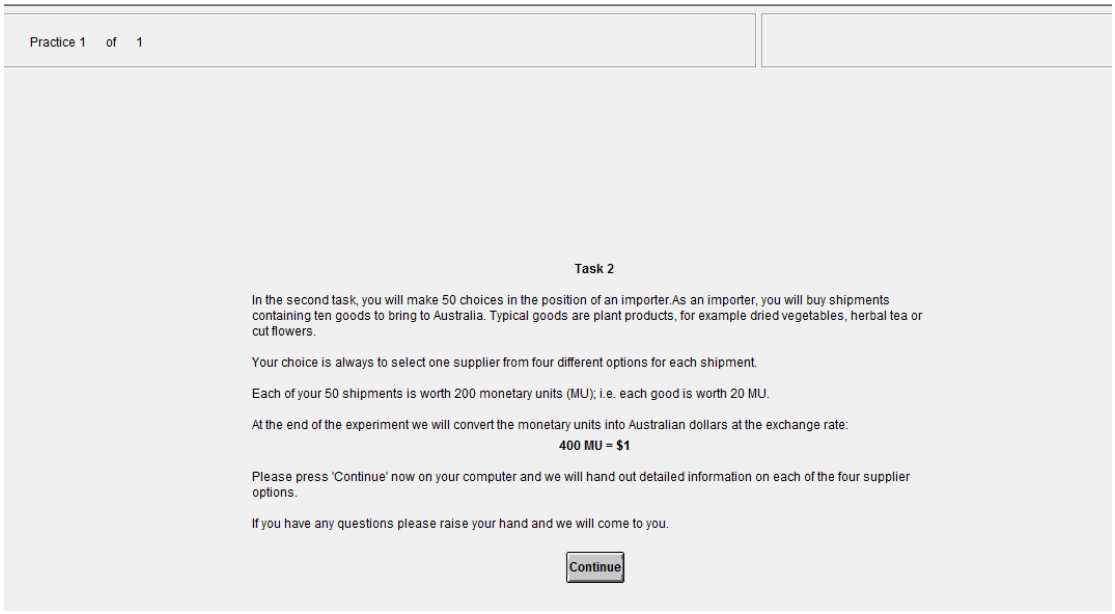


Figure 1: Computer-based instructions in z-Tree for the biosecurity inspection game

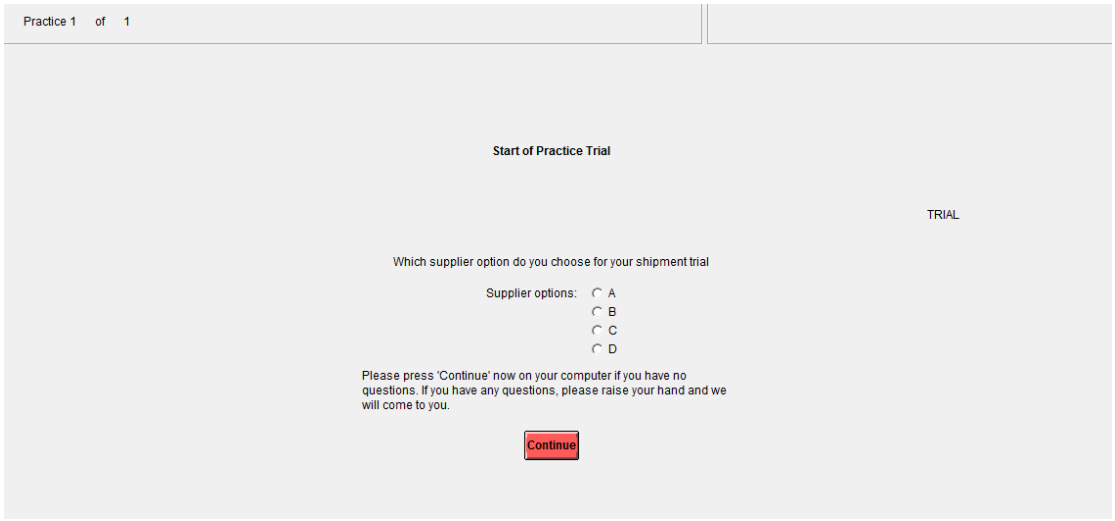


Figure 2: Computer screen in z-Tree when participants make supplier choices in the biosecurity inspection game

2.1.4 Using financial incentives to induce experimental characteristics

A key methodological innovation that encourages subjects to adopt prescribed roles in economics experiments is *induced value theory* (Smith, 1976). The basic premise is that, under certain sufficient conditions, properly applying a "reward structure to induce prescribed monetary value on actions" (Smith, 1976, 275) means subjects adopt the pre-specified characteristics of their experimental role, while "their own innate characteristics become irrelevant" (Friedman and Cassar, 2004, 26). Applying reward structures in this way appears natural in many economic interactions, since optimising behaviour and rationality lend themselves to "straightforward translations into experiments employing financial incentives" (Hertwig and Ortmann, 2001, 390).[5]

[5] A more extensive discussion of the role of monetary incentives in economics experiments is available in Chapter 11 of Guala (2005).

In our experiments, all tasks except the post-experiment questionnaire have monetary rewards attached to them. This is designed to impart similar earnings structures to the subjects as might be observed by importers in practice and encourage them to make considered choices. As an illustration of the type of rewards shown to subjects, Figure 3 shows the feedback given to participants after they have selected their supplier for a given round in the experiment.

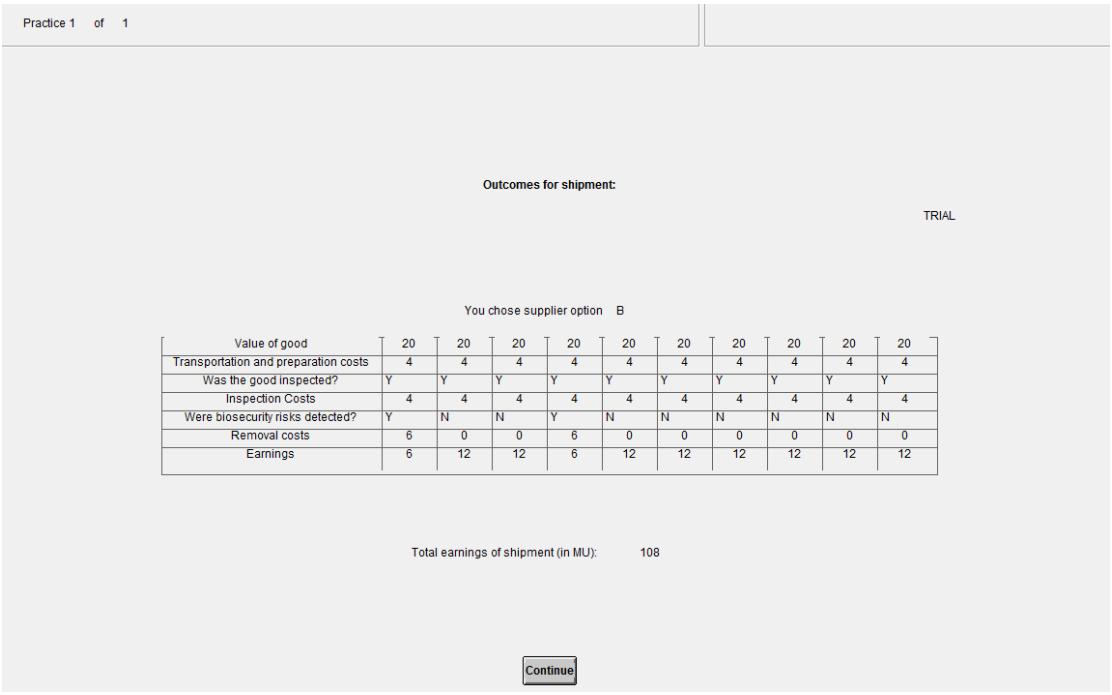


Figure 3: Participant feedback on the biosecurity inspection game with their earnings from their experimental choice

2.1.5 Randomisation to treatment and privacy in decision-making

In the experiments conducted as part of this project, two or four experimental treatments were being run as part of the one session, where the allocation to treatment was determined by the participant's seat number. While the subject recruitment software ORSEE[6] does not allow for complete randomisation of the subject pool between sessions,[7] within-session randomisation was achieved by participants drawing out their seating numbers from a lottery. This randomisation device ensured that, within each session, the experimental treatment received by a given subject was

6 ORSEE is the Online Recruitment System for Economic Experiments (Greiner, 2015) and is widely used by economics laboratories to recruit student participants.

7 The Monash Laboratory for Experimental Economics (MonLEE) maintains a pool of potential economics experiment subjects for which people (predominantly Monash University students) can sign up to. When a new experimental session is created, potential participants on this subject pool database are emailed through ORSEE and can sign up to any session with spots available. There are 24 computer terminals available in the laboratory, with enrolment in a session capped at 27 to allow for a few participants not turning up. If more than 24 people turned up to a session, people who turned up last would be paid the fee for showing up and not participate in the experiment. Recruitment requests are structured so that any individual can only sign up to one session of the experiment. There are penalties associated with failing to turn up to a session and procedures, through requirements to provide photographic identification, so someone cannot take the place of another person at a laboratory session.

not predetermined. Furthermore, participants did not know the treatments being run in a given session in advance, mitigating potential problems associated with selection into experimental sessions.

The researchers also took several steps to ensure privacy of individual-level decisions. During the experiment, any questions asked by participants were answered in private and no communication was allowed between experimental subjects. In addition, each subject received information and feedback only on his/ her own reward schedule and performance, with subjects paid in private in cash at the end of the experiment. As explained by Smith (2008), this procedure used to mitigate potential payoff externalities, since participants may experience negative or positive satisfaction from the rewards of others. If not adequately controlled for, this could reduce the experimenter's control over the induced reward structures applied in the experiment, potentially harming the replicability of experiments.

2.1.6 Using complementary experimental tasks to identify causal relationships

Laboratory experiments often require subjects to participate in several different, but related, tasks in the one session. Usually, there is one key experimental task of interest to the experimenter, with the other tasks constructed to provide additional information so the researcher can control for other influences on behaviour. This provides more confidence that differences in outcomes between pairwise-comparable treatments can be attributed causally to the factor that differs between the treatments, rather than an unmeasured, latent factor. For the experiments conducted in CEBRA Project 1404C, the three experimental tasks in addition to the biosecurity inspection game, namely:

- the abstract task to elicit the attitudes to risk of subjects, following Eckel and Grossman (2008);
- the post-experiment questionnaire to elicit other characteristics of the subjects, including attitudes to the environment and government interventions; and
- the paper-based incentivised task to assess how well the experimental subjects understood the inspection rules,

were all designed to ensure valid inferences could be made about the causal nature of differences in treatments and their effects on experimental outcomes. Figure 4 confirms the sequencing of each experimental session.

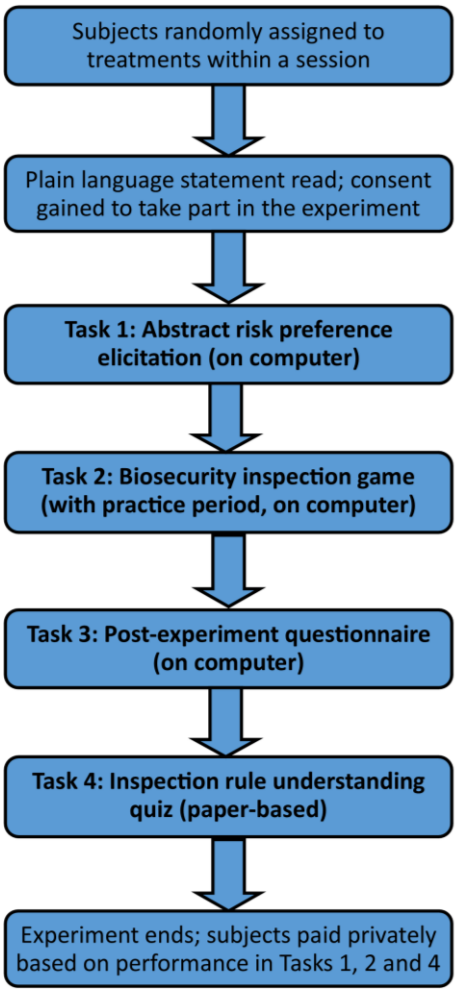


Figure 4: Schematic representation of the experimental session stages

2.2 Practical issues in experimental design

2.2.1 Role/s of the experimental subjects

As the purpose of the project related primarily to the response of importers to given inspection rules, the project team designed the experiment so that subjects in the main experimental task would take the role of importers. In this setting, the subjects made decisions about their suppliers in response to a predetermined set of rules imposed (or offered) in the experiment. This meant the computer took on the role as the biosecurity regulator in the strategic interaction, where the same rule applied for the duration of the experiment regardless of the experimental subject's supplier choices.

While the *individual-choice experiment* that results from this design is relatively easy to implement and the results straightforward analyse, it imposes restrictive assumptions about the behaviour of the regulator in this context. In particular, it implies that the regulator displays perfect commitment to a designated inspection rule for the duration of the experiment *regardless of the behaviour of the importer*.[8] For

[8] This assumption was also used by Rossiter and Hester (2017) to solve for the regulator's "optimal" choice of CSP rule parameters in their dynamic inspection game setup.

example, a biosecurity regulator might want to impose a version of "compliance hell" (akin to Greenberg, 1984) on an importer whose consignments routinely fail inspection to provide some incentive for them to improve their conduct. This type of arrangement is ruled out in the experimental design explored in this report, but would be possible under an experiment with social interaction between the regulator and importer.[9]

2.2.2 Framing the experiment's context

Economics experiments in the laboratory setting can either be naturally framed in the particular decision-making context of interest or can be subject to some level of abstraction, in which the context for conducting the experiment is expunged from the experiment (Loewenstein, 1999). Consultation with departmental officers and CEBRA colleagues indicated a strong preference for a naturally framed experiment, where the application to biosecurity inspection of plant-products would be clear to participants. Results from framed experiments generally allow them to be more easily understood by key organisational decision-makers, given the experimental context is more grounded in the reality of a particular policy application.

However, the natural framing of experimental instructions is not without potential problems. In particular, the context of the experiment may elicit particular behavioural patterns from the experimental subjects because it triggers latent psychological motivations from, for example, a personal or social norm.[10] While this may affect behaviour in the experimental setting, econometric techniques allow the experimenter to control for characteristics that might be associated with these underlying motivations. For this reason, the post-experiment questionnaire asks questions about subjects' attitudes to the environment, incursions of pests and diseases, government intervention to resolve environmental problems and political preferences. The effect of these factors on supplier choices is discussed in Chapter 6.

2.2.3 Adjustments to experimental language and explanations

Given the natural framing of the experimental instructions, an ideal situation would be for language used in the context of importing plant-based products to be incorporated into the instructions. However, care needed to be taken to ensure technical language in the experimental instructions does not impede subjects' understanding of the experimental tasks.

As many students in the Monash University Laboratory for Experimental Economics' (MonLEE's) subject pool have English as their second language, the project team took this into account by ensuring the instructions were clear and used simpler language where available. This was to ensure, to the fullest extent possible, that participants in the experimental sessions understood the tasks they are being asked to perform. For example, the term representing "consignments" was replaced by "goods" in the experimental instructions,[11] with "shipments" representing the 10 consecutive

---

9 This form of experimental design is discussed in more detail in Chapter 3.3.

10 See Elliott and Hayward (1998) for a more extensive discussion of the impact on choices using context-dependent framing of instructions.

11 The term "consignments" was inadvertently used in the CSP-3 algorithm diagrams provided to some participants as part of the paper-based task. Chapter 6.3.2 discusses the potential implications of this for the usefulness of that component of the experiment.

consignments for which the subject's choice of supplier would apply. The instructions also clearly explained the term "biosecurity risk material" and used examples to ensure participants understood what types of goods where included in "plant products".

Another aspect requiring careful explanation in the instructions was the nature of randomised inspections. Concepts from probability feature highly in this experiment, so it is necessary to explain these concepts to subjects without this background so they can still participate effectively. For example, when a shipment of 10 goods is inspected in monitoring mode under a CSP algorithm and the probability of each good being inspected is 20 per cent, the number of goods inspected should follow a binomial distribution with parameters 10 and 0.2.[12] This means that the number of goods actually inspected in a given shipment may be equal to, more than or fewer than two. This type of explanation was included in the instructions to avoid participants getting confused about how many goods they could expect to be inspected in a shipment.

2.2.4 Assessing subjects' understanding of the experimental instructions

The CSP algorithms used in the biosecurity inspection task are relatively complex and unfamiliar to the prospective subject pool. Experimental subjects' understanding of these rules may differ wildly, thus affecting how subjects may make supplier choices.

There are several approaches that can be used to assess, or at least partly control for, subjects' understanding of the experimental instructions, including:

- having subjects answer a quiz before beginning the main task to check their understanding of the instructions;
- using a complementary task that can provide additional information, independent of the main task, to assess their understanding of the CSP algorithm after the main experimental task; and
- asking respondents to report their level of understanding of the inspection rules, either before or after the main experimental task.

As part of this experiment, the project team agreed to use the second and third of the approaches listed above to control for subjects' understanding of the rules. The use of statistical controls for levels of understanding is likely to be more appropriate for applications in a policy environment. This reflects that there are likely to be varying degrees of comprehension of inspection rules and specific components of biosecurity assurance systems among stakeholders in the real-life decision environment – even among those highly familiar with the system.[13]

Using the complementary task in the laboratory allowed the project team to provide at least a partial assessment of whether presenting the CSP-3 algorithm using alternative visual stimuli (diagrams) may aid understanding of that rule's complex penalty

[12] This also assumes that inspections of each good are independent events and that an inspection failure was not detected within that shipment. Given the random number generator sitting behind the Q-ruler, this should be a good approximation for what happens in practice. Regardless, inspections are considered independent events for the purpose of this experiment.

[13] In the biosecurity context, this varying degree of understanding of how the CSP-3 algorithm operates was confirmed through interviews with green coffee bean importers and customs brokers as part of CEBRA Project 1304C.

structure without further adding to the number of experimental treatments. Subjects' self-reported understanding of the rules can also provide useful information to the experimenter, but relying on this solely is not without drawbacks. For example, participants may exhibit over-confidence and state that they comprehend the rule to a greater extent than their "true" level of understanding. Potential cognitive biases can then be evaluated through using an independent, objective measure of rule understanding as an adjunct to a self-reported measure.

The approach of having subjects answer a quiz to check their understanding of the instructions is standard in many experiments. However, in the experience of members of the project team, this pre-testing approach has been found to be problematic. In particular, getting past the testing phase can require significant intervention by the experimenter for some subjects in the laboratory session, thus de-valuing attempts to ensure all participants understand the rule. Given this practical concern and the project team's interest in investigating the effect rule understanding may have on performance in this environment, this approach to verifying experimental subjects' understanding of the environment was not pursued.

2.2.5 Number of choices participants are expected to make in the experiment

The decision around the number of choices participants make in an economics experiment has implications for the theoretical underpinnings of the experiment and may influence observed behaviours in environments with features akin to "cooperation". When seeking to investigate ongoing interactions, the experimenter must decide whether the duration of the experiment – that is, the number of choices or *periods* – is fixed or determined probabilistically, in addition to selecting the (expected) number of choices.

With a fixed number of periods in the experiment, the observed behaviour by subjects can be compared with predictions of a finite-horizon game.[14] Such a framing of the experimental task has the distinct drawback of inducing "end-game" effects in subjects' choices. The reward structures associated with the biosecurity game mean "good" behaviour now reaps benefits in the future from reduced inspection costs. As the number of periods left in the game reduces, the expected cumulative (future) payoff from choices entailing stronger compliance also gets smaller. This results in an incentive for subjects to "cheat" towards the end of the experimental task, in that it is in their interests to make choices consistent with short-term reward structures rather than those that deliver longer-term benefits that tend to be sustained by conduct more closely aligned with what the regulator would consider desirable.[15]

If an infinite-horizon game is the desired comparator, this type of interaction can be induced experimentally by using a random termination procedure initially developed by Roth and Murnighan (1978). In this approach, after each period, a random draw

[14] The payoffs to subjects in the finite-horizon strategic interaction could be undiscounted or subject to exponential discounting. In most cases, as is the case here, undiscounted payoffs are used when considering a finite-horizon interaction.

[15] According to Lugovskyy et al. (2017), the evidence suggesting reduced cooperation in experimental tasks with a commonly known fixed number of periods relative to those whose number of periods is probabilistically determined is mixed. In their own public goods game experiment, they find no consistent evidence of reduced cooperation over the course of the task but observe a more pronounced decrease in cooperation in the final round when the number of periods is fixed.

determines whether to conclude the game or continue for an additional period. When the probability of continuation is a fixed value, *p*, the procedure is equivalent to an infinite-horizon interaction where the discount factor attached to future per-period payoffs is *p*. Furthermore, as noted by Duffy and Fehr (2017), the expected number of future periods of the game from any period reached is *1/(1-p)*.[16]

From a policy perspective, the infinite-horizon representation is more appealing, as it avoids potential "end-game" effects which are a methodological artefact rather than something that would be observed in the regulatory environment. Despite this favourable property, the practicalities of implementing the experiment, as outlined below, led the project team towards designing the main task with a fixed number of periods. However, it could be an option for future research to consider designing an experiment around the infinite-horizon format of the game.

As noted earlier in the section, the CSP algorithms[17] already involve explaining concepts from probability to the experiment's participants. Adopting the random termination procedure would introduce more probabilistic notions into the experiment's structure, further complicating task instructions in an already challenging scenario. Relative to imposing a fixed number of participant choices, the added complexity associated with introducing the probabilistic ending of the game may not be worthwhile, as it may impede the ability of participants to fully comprehend the experimental task.

If the main experimental task were to be based on an infinite-horizon environment with discounting, then an appropriate discount factor (via the probability of continuation in the experiment) must also be considered. The choice of continuation probability has implications for the conduct of the experimental sessions and the number of choices made by subjects. Because importers must "wait" to receive the benefits of reduced inspections, after first building up a good history of compliance, a low continuation probability would be unlikely to sustain "cooperative" behaviour that would encourage choosing suppliers with lower approach rates. This tendency is amplified by the prospect that cost savings from avoiding inspection are relatively small – something that is likely to be a feature of the real-world biosecurity inspection environment for many pathways. A high discount factor also makes sense in the context of biosecurity inspections, since many of the larger importers who stand to benefit from the reduced rate of inspection will be those who bring in many (50 or more) consignments a year.

For an individual-choice task, as has been developed here, choosing a high continuation probability could create significant issues in a laboratory setting that seats 24 participants due to perceived inequities in the number of periods attained. Unlike other studies (e.g. Lugovskyy et al., 2017) with smaller group interactions and a lower continuation probability (of 0.8), the number of periods that people could play

[16] This "memoryless" property arises because the number of periods in the interaction follows a geometric distribution with success probability *p*.

[17] Unlike many economics experiments focused on infinite-horizon interactions, choices made by participants in this experiment are not "stationary replications"; instead, past choices can affect the payoffs available to participants for several periods, inducing "path dependence". Theoretically, this means the problem is not a "standard" repeated game, but a Markov decision process problem. From a practical perspective, solving this problem using simulation methods, whether undiscounted or discounted, would rely on some form of truncation of the infinite-horizon case back to a finite-horizon setting.

for does not differ too much and allows for some repeated play. In an individual-choice setting, having a higher continuation probability could result in significant boredom and/or envy among participants due to the disparity of the number of periods played, as suggested by the illustrative calculations in Table 2 below. Furthermore, the very high maximum number of periods played in a session with a more realistic continuation probability (0.95) means it would not be easy to repeat the experimental task within a session and obtain more data from those who made relatively few choices in the task.

Table 2: Illustration of different continuation probability and session sizes on length of sessions

Continuation probability	**0.8**		**0.9**		**0.95**	
For a given group/individual:						
Expected number of periods played	5		10		20	
Probability of game finishing in 10 or fewer periods (%)	89.3		65.1		40.1	
Probability of game finishing after more than 40 periods (%)	<0.01		1.5		12.8	
Number of groups/individuals in session	**8**	**24**	**8**	**24**	**8**	**24**
Expected maximum number of periods played in the session	12.7	17.4	26.3	36.3	53.5	74.1

Given the preference to design an experiment with a fixed number of periods and with no discounting, it remains to determine the number of periods subjects must play. The calibration analysis, discussed in more detail in Chapter 3, suggests that the differences in payoffs between alternative supplier strategies tends to be more substantial when the number of consignments imported is "large" (in the order of several hundred consignments), thus creating a long "chain" of choices. However, there needs to be care taken to balance the need for sufficient differentiation in payoffs between alternative supplier choice strategies and the possibility to induce boredom among participants, which may result in them making choices that suggest inertia.

A compromise position arrived at for these experiments was that subjects chose their suppliers for blocks of 10 consignments in a row. The choice of 10 aligns with the clearance number used for most of the experimental treatments and allows for an easier exposition of probabilistic concepts in the experiment. It also enables the experiments to mimic the potential for importers to engage suppliers on short-term contracts which may then be subsequently renewed (or not) depending on that supplier's performance.

Despite allowing for supplier choices to represent groups of 10 consignments, the calibration exercises suggested the number of supplier choices would still need to be reasonably large to obtain some differentiation in the payoffs. The end position agreed by the project team was that subjects would make 50 choices of their suppliers, where each choice represented 10 consignments. That meant the subjects took on the role of importers who bring 500 consignments of goods into Australia over the duration of the experiment.

2.2.6 Choice of metric for assessing experimental outcomes

The results for each experimental treatment are analysed in terms of changes in the average biosecurity risk material approach rate (henceforth the "implied approach rate") for the experiment's hypothetical plant-product pathway. This measure is of particular interest to the department, given it indicates the extent of non-compliance with biosecurity requirements. The implied approach rate can be calculated across experimental subjects (as shown in the plots over time in Chapter 5 of this report) or both experimental subjects and time periods (as shown in the Table 7 of the *Final Report*) for each treatment. It is a weighted average of the approach rates of the four supplier options shown in Table 1 of the *Final Report*, with the weights determined by the number of choices participants made of each supplier in the experiment.

The implied approach rate is only one criterion on which to base decisions about the "optimality" of biosecurity inspection rules. Furthermore, it is only a partial measure, since the regulator may also care about the costs incurred in undertaking inspections. As such, we briefly discuss two other potential metrics that could be used to assess the experimental outcomes.

One potential way to assess competing inspection rules is the leakage rate; that is, the (ex-post) rate at which biosecurity risk material enters the country following the inspection process. This criterion represents an assessment of the inspection process itself, because how well border inspections identify biosecurity risk material in consignments has more influence over this metric than importer behaviour. It also automatically favours inspection rules which feature greater intervention at the border. Such a methodology could be useful if the regulator's sole focus is on leakage, as opposed to the costs imposed on society through an inspection process. If the inspection process is assumed to be perfect, as has been done in this experiment (see Chapter 3.1 of this report), this criterion is not appropriate. This is because a mandatory inspection regime will *always* be superior *regardless of how importers respond to the rule*.

A cost-effectiveness measure based on the regulator's loss function, similar to that used in the game-theoretic model for biosecurity inspections developed by Rossiter and Hester (2017), is another criterion that could be used. In this representation, the regulator's loss function depends on two cost parameters, namely:

- the implied cost of biosecurity risk material leaking into the domestic economy because a contaminated good (in the language of the experiment) was not inspected; and
- the cost to the regulator of undertaking inspections of the goods.

When the inspection process is assumed to be perfect, the ratio of these costs and the behaviour of the importer matter for determining which rule is preferred by the biosecurity regulator. As mentioned earlier in this section, the individual-choice setup of this experimental investigation means we are silent on the regulator's objective function.

For inspection rules with the same form and parameters, the treatment comparisons under the implied approach rate and loss function criteria should align. However, the cost-effectiveness criterion is valuable to the regulator when seeking to compare rules with different rates of intervention (that is, average proportions of goods being inspected) at the border. Since most of our treatment comparisons assess rules with

the same rate of inspection, we have chosen to focus on the simpler implied approach rate metric. For illustrative purposes, we show how a loss function-related criterion could be used to compare inspection rules with different rates of intervention in Chapter 5.2 of this report.

2.3 Field experiments

Experimental economics techniques can also be applied in a more naturally occurring context through field experiments. These experiments blend aspects of experimental control with the lived reality of a situation, by taking the controls applied in laboratory experiments to the field and systematically relaxing some of these controls inherent in the artificial laboratory environment. In a sense, this bridges the gap between the laboratory setting and data collected through happenstance observation. This approach enables researchers to assess the behaviour of experienced subjects in the context relevant for applying government policy and, more generally, address research questions deemed hitherto difficult to answer using either laboratory-based experiments or data collected through natural observation (Levitt and List, 2009, 2).

While a field experiment constitutes part of the successor project (CEBRA Project 1608C), an understanding of the complexities and characteristics of field experiments is useful for placing the laboratory experiments used in this project in context of the broader research and policy agenda. To that end, Harrison and List (2004) propose six factors that can be used to determine the field context of an experiment:

1. the nature of the subject pool. While laboratory experiments would typically draw on the undergraduate student population, field experiments would seek to draw subjects to represent the target population in the economy (for example, importers of plant-based products) or the general population;
2. the nature of the information that the subjects bring to the task. In contrast to artificial laboratory constructs, field subjects bring knowledge of the existing institutions and experience with tasks related to the experiment, which may affect their behaviour in practice;
3. the nature of the commodity. Using physical goods or actual services in the field may result in different behaviour from abstractly defined laboratory experiment constructs;
4. the nature of the task or trading rules applied. Field experience could result in different *heuristics* being applied in decision-making compared with experimental subjects inexperienced in the task. Another aspect of this relates to the parameters chosen in any laboratory experiment, and how they relate to those experienced in the field;
5. the nature of the stakes. The larger consequences of actions in the field compared with the laboratory may have an influence on behaviour; and
6. the nature of the environment that the subject operates in. Different strategies or heuristics may appear in the natural decision-making context that are not admitted in the more abstract laboratory setting.

These characteristics give rise to an (admittedly incomplete) classification structure for field experiments (Harrison and List, 2004, 1014), which go beyond the conventional laboratory experiments.

- *Artefactual field experiment:* As for a conventional laboratory experiment, but with a nonstandard (that is, non-student) subject pool.
- *Framed field experiment:* As for an artefactual field experiment, but with field context provided by the commodity used, the tasks undertaken or the information available for subjects to draw upon.
- *Natural field experiment:* As for a framed field experiment but which occurs in the natural environment of the tasks and where subjects may not know they are participating in an experiment.[18] This is the type of field experiment proposed to be conducted as part of CEBRA Project 1608C.

2.4 Experiments as a test-bed for designing institutions

2.4.1 The economic design process

This project forms part of a multi-stage process to discover the policy settings (including rules, incentive structures and monitoring practices) that align the actions of import-supply chain participants with the primary government objective of maintaining Australia's high biosecurity status. The ultimate goal of these projects is to design new (or refine existing) protocols and regulations governing the process of importing goods of potential biosecurity concern into Australia.

Figure 5 summarises the methodology for the economic design. This sequencing, and the use of laboratory experiments in general, has become established in the economic design process (see, for example, Roth, 2002; Plott, 1987). This is part of well-articulated and established form of economic inquiry.

CEBRA Project 1304C sought to assess incentive, information and implementation issues in the current biosecurity assurance system and develop new candidate protocols drawing upon microeconomic theory, quantitative administrative data and qualitative evidence from stakeholder discussions. This constituted the "problem diagnosis" and "policy design" phases of Figure 5 above and allowed candidate protocols and procedures to be identified.

This project, CEBRA Project 1404C, implemented the phase "test and refine the policy design". This phase employs well-developed methodologies in experimental economics to examine key aspects of potential protocols in a controlled (laboratory) environment that allows for structured observation and the identification of causal linkages. This carefully crafted experimental setup allowed candidate protocols to be assessed for their robustness and applicability for the natural environment using a "test-bed" or "wind tunnel"-like approach (Plott, 1997).[19] As part of the experimental design phase, we also used computational economics to simulate importer behaviour and assist understanding of the incentive structures facing importers in their decision-making.

---

[18] The issue of obtaining informed consent for human subject research is a complex issue, governed by conventions such as those of the National Health and Medical Research Council, Australian Research Council and the Australian Vice-Chancellors' Committee (2015). For more discussion in the field experiment context, see List (2008) and Levitt and List (2009).

[19] Similar processes have also been used successfully in the business context; see, for example, Chen (2005).

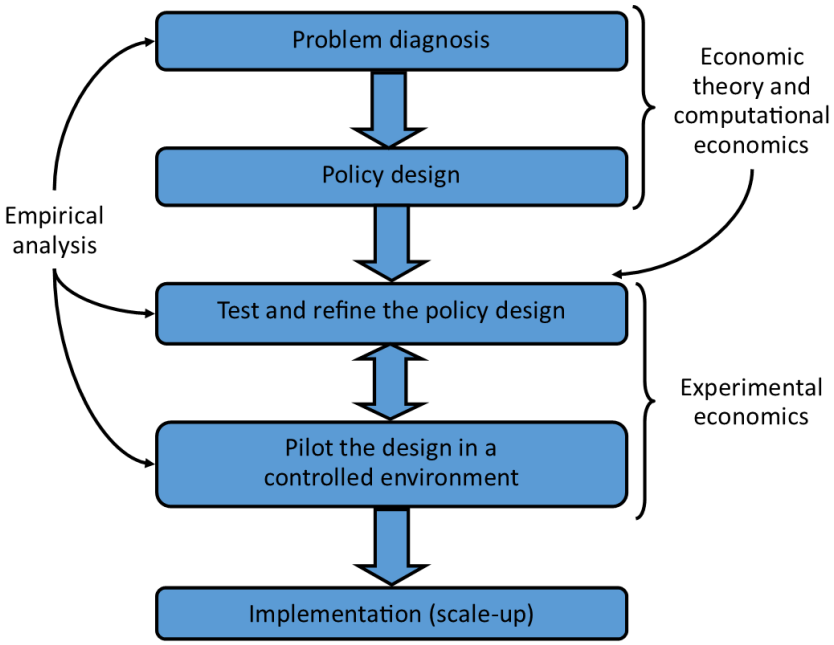


Figure 5: Economic design methodology schematic representation

The stage "pilot the design in a controlled environment", to be covered in CEBRA Project 1608C, will take protocols, refined through the test-bed process, to real-world pilots on a few select plant-product pathways. The environment will still be subject to some controls, though to a lesser extent than is possible in the laboratory. This subsequent project should enhance the evidence base for the department to support the broader implementation of new regulatory settings that are practical, implementable, cost-effective and assist the Australian Government achieve its policy objectives in biosecurity.

2.4.2 Laboratory experiments in the design refinement phase

Laboratory experiments have been used as a precursor to rolling out new designed policies and procedures, with several applications dealing with the design of auctions,[20] smart markets (for example, Brewer and Plott, 2003) and matching markets.[21] In the regulatory sphere, laboratory experiments have been used to understand regulatory decision-making by committee (Grether et al., 1981), inform the design of incentive regulation schemes (Cox and Isaac, 1987) and the effects of market practices on competition (Grether and Plott, 1984).

Part of the appeal for using laboratory experiments the design process is they are a safe, low-cost environment for testing policy mechanisms based on actual human behaviour in response to the experimental rules. In the test-bed approach, laboratory experiments can examine how rules operate in simple cases based on "a simple working prototype of a process that is going to be employed in a complex

[20] Notable examples include auctions for telecommunications licences (for example, Plott, 1997; Goeree et al., 2006) and emissions trading systems (Cason and Plott, 1996) in the United States, and native vegetation offsets (Nemes et al., 2008) in Australia.

[21] See Roth (2013) and the references therein for an extensive discussion of examples of matching markets, including medical labour markets, models of school choice and kidney transplants.

environment" (Plott, 1997, 607). This is primarily so "the nature of any problems detected can be identified and studied" (Plott, 1997, 607).

Laboratory experiments of this kind can also be a way for policymakers to reduce implementation risk. This is because these techniques help researchers discover practical problems with the proposed rules or mechanisms relatively quickly, including problems that could prove very costly if they surfaced during practical implementation. As Plott (1997) explains, if a particular mechanism fails to perform well during experimental evaluation and its predictions do not align well with theoretical predictions, there may be little cause to believe it could work well (or as predicted) in the real-world application of interest.[22] In effect, this type of framework allows potential policy flaws or "design bugs" to be identified as part of a failure of proof of concept.

Ideally, laboratory experiments could be a precursor to a small-scale field pilot to "permit sharper and more convincing inference" (Harrison and List, 2004, 1009).[23] In this case, laboratory experiments could be used to develop preliminary estimates of reactions to changes in rules or policy settings in an environment that strips out extraneous factors. These more highly controlled experiments may also provide an indication as to what types of rules should be the focus of the field pilot. Similarly, laboratory and field experiments could be used in a complementary way, such as to explore potential confounding aspects or where laboratory experiments could be used to investigate aspects that may be difficult to investigate in the field.[24]

2.5 Translating experimental findings to policy practice

There has been considerably attention placed on the use of experimental methods in economics, and in the social sciences more generally, particularly around notions of whether particular conclusions or inferences in experimental settings adequately reflect the "truth" (Roe and Just, 2009, 1266). As noted by Jiminez-Buedo and Miller (2010), the focus has been on two concepts relating to the worth of experiments, namely:

- *internal validity*, around whether a method allows for the appropriate attribution of causal relationships, or lack thereof, between two variables; and
- *external validity*, which relates to the generalisability of presumed causal relationships across different measures of the underlying characteristics, different subject pools or different contexts.

Some researchers posit a trade-off between different notions of validity for alternative research designs. For instance, Roe and Just (2009) contend there is a research

[22] This idea of stress-testing mechanisms before they are put to use in public policy should also deal with a mechanism's robustness to various forms of misbehaviour and gaming. See Bolton (2013) for further discussion of this important issue.

[23] Critically, laboratory and field experiments should not be seen as an "either-or" proposition, but as two research methods that can help address different components of interconnected research questions and gradually accumulate robust knowledge about causal relationships.

[24] Roe and Just (2009) note that the type of interventions and/or range of stimulus can be restricted in field experiments. For biosecurity inspection arrangements, ethical or other commercial interest reasons preclude some types of interventions from being tested in the field. As a result, the project team's approach has been to assess these types of interventions in the laboratory so as to inform the field trial's design.

methodology spectrum "defined by the degree of verifiable exogenous variation within the economic context that produces the data" (Roe and Just, 2009, 1267). Laboratory experiments (with their relatively high internal validity and relatively low external validity) comprise one end of the spectrum, with the other being naturally occurring field or market data (which has relatively low internal validity but relatively high external validity). Field experiments, they argue, lie somewhere in the middle of the spectrum and can offer medium to high level internal and external validity.[25]

Others in the literature adopt a more nuanced approach to issues of external and internal validity.[26] Jiminez-Buedo and Miller (2010), for example, contend that replicating experiments with slight variations to account for potential confounding effects that may in part be responsible for observed behaviour in the real-world decision-making context can help secure both internal and external validity. They proceed to argue that, rather than a trade-off between internal and external validity, the perception of low external validity arises because of over-confidence displayed in seeking to extrapolate causal relationships established experimentally to other settings.

By construction, carefully crafted laboratory experiments, where potential confounding elements have been adequately controlled for or eliminated, should allow researchers to attribute differences in outcomes to treatments in a causal manner, admitting internal validity. In contrast, the internal validity of field experiments may be affected by aspects such as systematic differences in treatment groups or the presence of uncontrolled variation in unobserved variables.

Where laboratory experiments can suffer is in terms of the ability to generalise their conclusions beyond the experimental setting. For instance, carefully designed experiments will not necessarily allow researchers to understand the prerequisite conditions under which an experimentally established causal link may be observed in a given real-life context (Jiminez-Buedo and Miller, 2010, 319). Instead, a range of intervening factors, some of which may be difficult to measure objectively, may mask the ability to confirm the relationship in environments subject to more limited control by the researcher. The nature of the subject pool,[27] more limited variation in the levels of stimulus[28] and the lack of contextual elements organic in the real-world policy

---

25 Roe and Just (2009) also note that combining several approaches – such as laboratory and field experiments – to study the same phenomenon can reduce tensions between internal and external validity.

26 Jiminez-Buedo and Miller (2010) acknowledge another methodological position espoused in the literature is that internal validity is a prerequisite for an experiment to be able to say anything about the external environment.

27 For example, Koudstaal et al. (2016) provides evidence that entrepreneurs have different attitudes to making losses than managers or employees, even though their attitudes to risk and ambiguity do not seem to differ markedly. Some in the literature (for example, Harrison et al., 2009) even argue that students recruited to conventional laboratory experiments have different risk preferences relative to the potential subject pool of all students as a result of commonly used experimental procedures. This view is not universally supported; see, for instance, Cleave et al. (2013).

28 Evidence on the strength of this effect is mixed. The meta-analysis of Camerer and Hogarth (1999) suggests that "higher levels of incentives have the largest effects in judgment and decision tasks" but "in games, auctions, and risky choices, the most typical result is that incentives do not affect mean performance, but incentives often reduce variance in responses" (Camerer and Hogarth, 1999, 34).

environment, as discussed in Chapter 2.3, inhibit the ability to extrapolate experimental results in a straightforward way to the field.

While some in the literature (for example, Charness and Fehr, 2015; Herbst and Mas, 2015) argue that *quantitative* predictions can carry over from the experimental laboratory to the field context, this assumption is probably too strong in most situations for the reasons described above. In preference, we consider the stance of Kessler and Vesterlund (2015) more instructive for the use of experiments for policy purposes, in that:

- laboratory experiments can help uncover *principles of behaviour*, which are themselves general and externally valid;
- *qualitative* findings around the *direction* of treatment effects from laboratory experiments *should* be generalisable; and
- the use of simplifying assumptions to enable internal validity for laboratory experiments implies the *magnitude* of observed (treatment) effects *will likely differ* between the controlled laboratory environment and other environments.

These principles provide no guarantee of being able to extrapolate from the highly controlled laboratory to less controlled policy environment; however, they are useful for framing how to consider experimental results as part of the policy development and reform process.

In the context of designing and implementing biosecurity inspection frameworks, this experimental study has been designed to seek insights into *general economic behaviours* in response to different types of inspection rules. In several cases, the experimental treatments take the form a cross-check to ensure elements of the inspection protocols that could be adopted in the field appear to work in the direction (but not necessarily magnitude) that is expected based on economic theory. As induced value theory allows the experimenter to mimic the types of incentive structures inherent in these inspection rules, findings about responses to incentives in the experiment are likely to transcend the subject pool (university students) and apply to a large degree to the specific target population, namely importers of plant-based products. Furthermore, because of the way in which alternative implementation strategies mirror the natural context, it could also be expected that responses to various behavioural devices are likely to carry over to the importer population of interest. Whether the direction and/or magnitude of experimentally determined effects *actually* carry across to the natural regulatory environment can only be determined through careful field work.

Our discussion of experimental results, particularly in Chapters 5 and 6 of this report, follows the three principles articulated above by largely focusing on the direction of treatment effects rather than emphasising magnitudes.

3. Translating the biosecurity inspection game into an experiment

In designing experiments, one of the critical requirements is to understand the key influences expected to affect behaviour in the public policy setting. The interactions between importers and the regulator for biosecurity inspections are complex, so there needs to be a degree of simplification associated with translating real-world interactions into the experimental laboratory. As discussed in earlier chapters, it is also not feasible (or worthwhile) to test every potential aspect affecting decision-making in the public policy application. Simplifications then allow specific aspects of the framework that are of interest, or considered most "vulnerable" in some sense, to be analysed and their effect on behaviour isolated.

Theoretical frameworks may provide a useful starting point for identifying the factors likely to be influential on behaviour. In a sense, the laboratory experiments involve testing normative predictions of behaviour (that is, how agents should respond) based on stylised representations which impose strong assumptions on the way in which agents make decisions.

In this chapter, we review the factors influencing decision-making by the regulator and importers in the biosecurity inspection context. This involves considering an alternative version of the model examined in Rossiter and Hester (2017) where the importers have a choice over their supplier, rather than forming a vertically integrated supply chain. We do this primarily to simplify the choice environment for the experiment, though this comes at the cost of making further allowances to avoid other factors confounding the interpretation of the experimental results.

This chapter also discusses some of the practical considerations taken into account in designing these experiments. We also consider how other factors, such as attitudes to risk, that could be important to the experimental results can be controlled for in this context. The important issue of how to calibrate the parameters chosen for the experiment and generate theoretical predictions of behaviour is considered at the end of this chapter, together with discussion of some important policy implications identified through this exercise.

3.1 Choices and factors influencing decision-making for the regulator

Even under a simplified decision environment, there are many choices the biosecurity regulator can make which can affect importer behaviour. In broad terms, these include:

- the form of inspection rule applied, which for the continuous sampling plan (CSP) family of rules dictates the penalty structure for failing an inspection when in monitoring mode. For more detail on the structure of the CSP rules, see the box below;
- the key rule parameters, namely the clearance number, monitoring fraction and, for the CSP-3 algorithm, the tight census number (usually set at four);
- the level of information given to the regulated entities around the specifics of the inspection rule implemented; and
- the amount and nature of feedback on a regulated entity's performance under the inspection rule.

Continuous sampling plan algorithms

In this box, we introduce the three continuous sampling plan (CSP) algorithms considered in previous studies for implementation by the department. The most basic of the CSP family rules is the CSP-1 algorithm, which was introduced in Dodge (1943) and is illustrated in Figure 6.

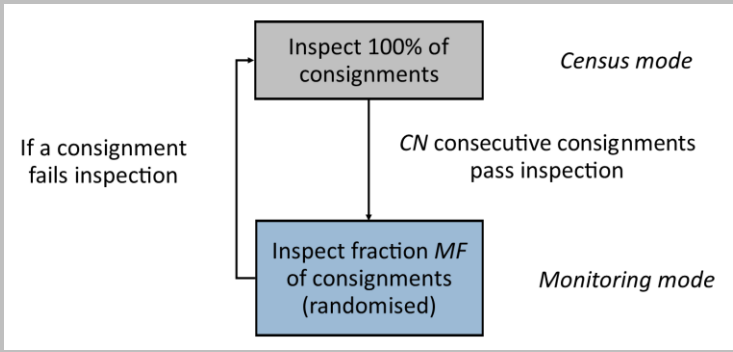


Figure 6: Schematic representation of the CSP-1 algorithm

When a new importer starts on this algorithm, they are usually subject to mandatory inspections (in "census mode") until they build up a good compliance record. Two key parameters for the regulator to choose in this rule are:

- the clearance number (*CN*) – the number of successive consignments that must pass inspection for the importer to be eligible for a reduced inspection frequency; and
- the monitoring fraction (*MF*) – the reduced inspection frequency and probability that a given consignment is inspected in "monitoring mode".

If an importer's consignment fails inspection when the importer is in "monitoring mode", their subsequent consignments are subject to mandatory inspection in "census" mode. The importer only receives the reduced inspection frequency again after another *CN* successive consignments pass inspection.

The CSP-2 and CSP-3 rules documented in Dodge and Torrey (1951) have less severe consequences for occasional non-compliance when an importer is on the reduced inspection frequency *MF* relative to the CSP-1 rule.

In the CSP-2 algorithm (Figure 7), if an importer's consignment fails inspection in monitoring mode, then they continue to be inspected at the reduced rate (*MF*) while the regulator keeps track of the number of inspections passed since the last recorded failure. This part of the algorithm is usually referred to as "failure detection mode". Provided the importer passes inspection *CN* times since their last failure, they remain eligible to be inspected at the reduced rate of inspection; otherwise, on recording another failure within *CN* consignments of the previous one, the importer's consignments revert to mandatory inspection until they pass inspection *CN* times in a row. Intuitively, this provides less of a "cost" to the importer if recording a failure in one inspection does not increase the probability that future consignments will be more likely to fail.

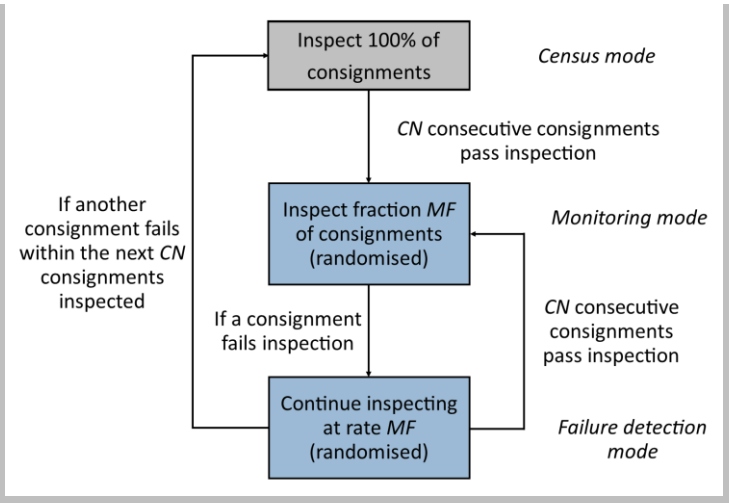


Figure 7: Schematic representation of the CSP-2 algorithm

The CSP-3 algorithm, shown in Figure 8,[29] adds another layer of complexity to the CSP-2 algorithm. This is designed to provide extra protection to the regulator against a sudden systematic problem that would significantly raise the likelihood of a consignment failing inspection. It does this by making the next four consignments following a failure subject to mandatory inspection in what is referred to as "tight census mode". The other features of the CSP-2 algorithm, such as ignoring past failures if they occurred more than *CN* inspections ago, are retained by the CSP-3 algorithm.

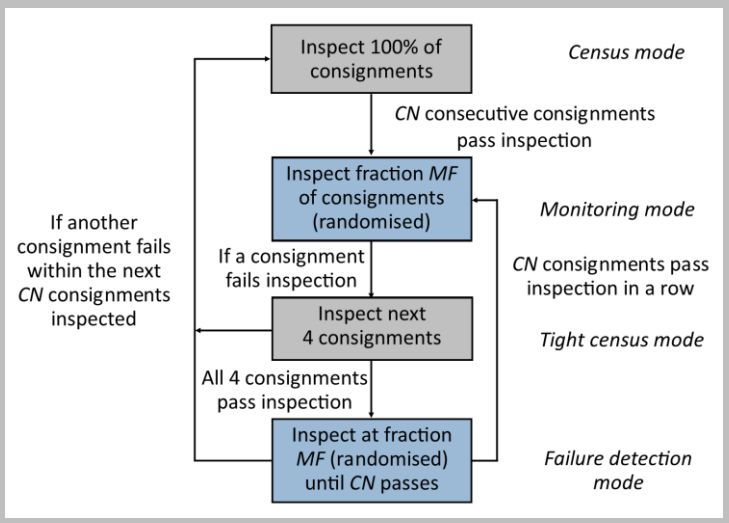


Figure 8: Schematic representation of the CSP-3 algorithm

Table 3 provides a more comprehensive account of these parameters and a brief description of how these are applied in the experimental setting.

[29] The version of the rule used in this paper follows the practical simplification suggested by Robinson et al. (2012).

Table 3: Choice variables available to the biosecurity regulator

Choice dimension	Options/scope	Comments	Use in the experiments
Type of rule adopted	CSP-1 CSP-2 CSP-3	The rules differ according to the penalty they place on failing inspection when in "monitoring" phase. The CSP-1 rule provides the harshest penalty, by requiring the importer to return to the mandatory inspection phase. The CSP-2 rule provides the greatest leeway by allowing the reduced inspection rate to continue provided another failure is not detected within a given number of inspections. The CSP-3 algorithm's penalty is somewhere between these rules. The different degrees of complexity in the consequences of failing inspection also affect how easily these rules can be explained to and understood by importers.	Most treatments will use the CSP-1 algorithm in line with the recommendations in Rossiter and Hester (2017). The CSP-3 algorithm will be used in some treatments to compare the results with current practice. The CSP-2 algorithm was found to be the least preferable for the regulator and will not form part of the experimental treatments explored.
Clearance number (CN)	Integer value (usually 5 or larger)	A higher clearance number requires the importer to demonstrate greater compliance with biosecurity requirements before having access to the benefit of a lower frequency of inspection. A higher value also raises the cost of failing inspection when in the "monitoring" phase.	In most treatments, this will be fixed at 10, in line with the largest clearance number currently used by the department by commodities on the Compliance-Based Inspection System (CBIS).
Monitoring fraction (MF)	Probability (strictly) between zero and one	A lower monitoring fraction benefits the importer, as they avoid costs associated with making consignments available for inspection. However, a low monitoring fraction may result in an importer acting to increase the approach rate of biosecurity risk material in consignments during the monitoring phase. This could occur if the amount of abatement effort was reduced or the importer chose lower-cost suppliers whose consignments are more likely to contain biosecurity risk material.	In most treatments, this will be fixed at a value between 0.1 and 0.5, reflecting the current range of values used by the department.
Tight census number (TC)	Integer value (set at 4 in the original algorithm)	The number of mandatory inspections in a row under the CSP-3 algorithm after failing inspection in monitoring mode.	This will be set at 4 for the CSP-3 treatments.

Table 3 (continued): Choice variables available to the biosecurity regulator

Choice dimension	Options/scope	Comments	Use in the experiments
Basis for rule application	Importer, supplier, producer or a combination of these	The Compliance-Based Inspection System (CBIS) uses the CSP-3 algorithm on plant-product pathways and is currently applied based on an importer's inspection history; in contrast, the Imported Food Inspection Scheme (IFIS) uses the supplier's compliance history to determine the inspection frequency. Suppliers are predominantly responsible for mitigating biosecurity risks in production, while importers may have the option to choose lower-risk suppliers from which to source their product. This only makes a material difference to decision-making for importers who are not vertically integrated.	This will be fixed across treatments as the importer being the basis for applying the inspection rule, in line with current practice under CBIS. The experimental structure expressly assumes a strong link between supplier choice and the likelihood of complying with biosecurity requirements. Adding a further layer of complexity would complicate the importer's decision space.
Information provided to the importer on the inspection rule	Rule parameters Details of operation	The inspector could choose to fully describe the rule and its parameters, including discussing the consequences of failing an inspection. At present under CBIS, the department discloses the clearance number but not the monitoring fraction; instead, the monitoring fraction is described as being between 10 and 50 per cent. It also doesn't completely specify the consequences of failing an inspection.	This will be varied between treatments. For the CSP-1 treatments, the rule will be fully specified or have the monitoring fraction within a range. For the CSP-3 treatments, the rule will be full specified, have the monitoring fraction within a range or both the monitoring fraction and tight census number described in a "vague" manner.
Information provided to the importer on their performance under the rule	Number of avoided inspections Failure rates Inspection outcomes relative to others on pathway	Reporting the regulator's administrative data could encourage greater understanding of the consequences of an importer's behaviour. Making comparisons relative to others on the pathway (e.g. relative to the pathway average) could encourage importers to change behaviour to gain a commercial advantage if the rewards are sufficiently high. This aspect of the inspection protocols is one to be tested explicitly in the experiments.	This will be varied between treatments based on the importer's own performance. Gain and loss frames will be used to highlight particular benefits (or costs) of the supplier choices made by subjects.

In the context of this experiment, we are not expressly testing the impact of changing the clearance number and monitoring fraction parameter values in the CSP family of rules on importer behaviour. While this is an important aspect worthy of examination, the implications of changing these parameter values in isolation are clear from a theoretical perspective; these are discussed, where relevant, in Table 3. Given the large number of dimensions of interest in this context, the project team decided that experimental verification of this was less critical than other aspects of these inspection rules.

The regulator's decisions around the choices it can make on a given pathway are influenced by a number of other parameters that affect its payoff from choosing particular actions. These parameters are discussed in Table 4.

Table 4: Costs parameters and decision-error probabilities influencing regulator choices

Parameter	Comments
Implicit cost of carrying out an inspection	Includes the opportunity cost of government spending on inspection activities, imposts which are not fully cost-recovered and accounts for any "referred" dislike for regulatory burden from the government's perspective. All other things held constant, a higher implicit cost of inspection would mean that the inspector would prefer to inspect less frequently.
Cost of biosecurity risk material leaking	This tries to capture a weighted average of the costs of incursions for domestic production, plus any eradication or rectification activities. If leakage has higher consequences, then higher rates of intervention will be preferred. In some cases, the very high consequence may mean mandatory inspection is the best way of dealing with potential biosecurity risks.
Probability of detecting hazards when present in a consignment subject to inspection	Measures inspection efficiency and measures one of the types of decision errors involved in inspection. A lower value of this parameter results in greater leakage of biosecurity risk material into the importing market. This parameter is partly under the influence of the inspector, as this parameter could be affected by training. A value of one for this parameter indicates perfect decision-making in inspections according to this dimension.
Cost of unnecessarily delaying consignments when hazards not present	Incurred when error made in inspection processes. Mainly consists of "referred" regulatory burden on importers and/or opportunity costs of resources involved in completing treatments that were not required. Thought to be of much lower consequence to the inspector than the cost of biosecurity risk material leakage.
Probability of falsely detecting hazards when they are in fact not present	Represents the second type of decision error in inspections. This parameter is partly under the influence of the inspector, as this parameter could be affected by training of frontline staff. A value of zero for this parameter indicates perfect decision-making in inspections according to this dimension.

For the experiments, the cost parameters in Table 4 will be treated as fixed, since the focus is on assessing importer behaviour in response to given rules. Of course, in deciding to use compliance-based inspection protocols on a particular pathway, the department would need to be comfortable with the possibility that biosecurity risk material could leak into the environment. From a theoretical perspective, this implies

that the cost of leakage cannot be "too large" relative to the cost to the inspector of carrying out an inspection.

Imperfect decision-making in inspection process was emphasised as an important aspect affecting importer decisions around biosecurity risk mitigation effort in Rossiter and Hester (2017). However, from the experiment's perspective, introducing the potential for errors in regulator decision-making would complicate the information provided to subjects and not necessarily result in findings that could easily be translated to the field. Adding that extra layer of complexity may also make it more difficult to identify causal links arising from the experimental data. For these reasons, the experiments will assume that the inspection process is "perfect", in that the inspection decision accurately reflects whether or not a consignment contains biosecurity risk material.

In practice, the costs an importer faces in the inspection process and incurred for the treatment of consignments containing biosecurity risk material is under the partial control of the regulator. This dependence is discussed for completeness in Table 5.

Table 5: Parameters affecting importer strategy under partial control of the regulator

Parameter	Comments
Cost to the importer of a physical inspection	The government has control over the direct charges set for inspection, while its system rules around the inspection process can also affect the delays incurred by importers. While this is assumed to be constant regardless of the importer's performance history, decisions around priority queueing for importers with a good compliance history could reduce the cost of inspection for those "good" importers.
Cost to the importer of rectifying consignments with biosecurity hazards	The government has some control over this cost through its ability to require particular treatments or restrict the available set of treatments that importers can choose from.

3.2 Choices and factors influencing decision-making for the importer

3.2.1 Framing options in the experiment for importer choices

Analysis of pathway data as part of CEBRA Project 1304C suggested that importers tend to fall into two broad categories, namely:

- those that are, or act as if they are, vertically integrated. For example, this could be through arrangements such as being the Australian distribution arm of a multinational business; and
- those that have freedom to choose their suppliers and obtain their products from a wide range of sources.

These two types of supply-chain structures result in a very different set of choices available to importers to change the costs they incur as part of the biosecurity inspection process. The different actions available and the costs faced by these two stylised types of importers need to be taken into account in designing compliance-based inspection protocols. Ideally, the protocols chosen by the

department should work effectively to support the Australian Government's biosecurity objective for both types of importers.

From the experimental perspective, if similar lessons can be learned from both supply-chain structures, then it is preferable to focus on the structure that admits a simpler choice set for experimental subjects and potentially yields clearer predictions of behaviour. Some of the key elements of interest to assess include:

- the ability for designed inspection protocols to encourage importers to reduce the biosecurity risk material approach rate of their consignments;
- the circumstances under which protocols may encourage behaviours that raise the likelihood of biosecurity risk material being present; and
- the influence of inspection rule parameters in encouraging different behaviours; and
- the dependence of appropriate inspection rule parameters on private information of importers, including their costs of being inspected, treatment costs and costs associated with changing their behaviour.

As discussed in Rossiter and Hester (2017), vertically integrated importers would make choices over the level of "fixed" technology to use and then, for each consignment, choose the level of effort per consignment. The influence of these choices on the probability of biosecurity risk material being present in a consignment is conveyed through an abatement technology function. The types of choices able to be made by the importer interact through this function and lead to higher complexity in the importer's action space.

On the other hand, a simpler choice framework for the importer is available if they are able to select their supplier. This type of importer is less likely to be able to influence the production processes of a supplier directly;[30] at best, the importer may be able to provide the supplier with information on inspection outcomes. As Australia is a relatively small export destination, overseas suppliers are less likely to make adaptations specifically for the Australian market. In this way, an Australian importer could be provided "take it or leave it" offers from overseas suppliers and is forced to choose from the options available. This also implies importers would be unable, or at best have limited ability, to influence the biosecurity risk material approach rate for a given supplier's consignments.

Although importers tend to source their products from multiple suppliers, evidence from stakeholder interviews suggested this was driven by considerations other than biosecurity, such as satisfying particular customer requirements or to ensure continuity of supply. As biosecurity-related factors are being assessed in these experiments, the choice set can be simplified so that importers can choose one supplier at a time for a given consignment.

Provided the various inputs affecting importer choices can be appropriately calibrated, the model with supplier choice is able to demonstrate similar patterns of predicted behaviour as the vertically integrated importer model for the elements of interest in designing biosecurity inspection protocols. Given the simpler choice set available for the supplier-choice model, this was the approach adopted in these experiments.

[30] The observations made about this type of import-supply chain structure were supported by intelligence gleaned from discussions with stakeholders across a range of plant-product pathways as part of CEBRA Project 1304C. See Chapter 5 of Rossiter et al. (2016) for more details.

3.2.2 Theoretical model parameters affecting importer choices

Table 6 below highlights the factors influencing the decisions made by importers with choice of supplier in this theoretical framework.

Table 6: Parameters influencing importer choices when they have choice of supplier

Parameter	Comments	Use in the experiments
Landed price of consignments (for a supplier)	Accounts for all non-biosecurity-related costs of purchasing the imported consignments for resale in the domestic market. From a practical perspective, this would include transportation costs but exclude any costs associated with biosecurity inspection process.	Among the choices of suppliers in the experiment, these two quantities are assumed to be negatively related. This helps rule out supplier choices which would be expected to be "dominated" by other choices. The "schedule" of these two quantities are the same for all experimental treatments.
Probability of biosecurity risk material being present in a consignment (for a supplier)	This could represent a single point (i.e. if the probability was known *a priori*) or a probability distribution that reflected the importer's knowledge (or belief) of the likelihood of biosecurity risk material being present in consignments. Beliefs could be informed by a supplier's compliance history and any information provided by the regulator to importers.	
Process for updating beliefs around supplier biosecurity compliance	This describes how importers revise their beliefs about how "clean" a supplier is based on evidence from inspections. For example, this could be through Bayesian updating of their prior beliefs around a supplier's compliance.	Approach rates are given to experimental subjects so that beliefs and belief formation processes regarding approach rates do not affect behaviour in the experiment.
Cost of changing to another supplier	Reflects both monetary and non-monetary costs associated with an importer deciding to change supplier. These include the time taken for the importer to learn about the supplier's operations and the time required by the supplier to understand an importer's specific needs. Higher switching costs would be expected to reduce the extent to which switching suppliers is observed.	Switching costs are taken to be zero for the context of the experiment. This implies potentially greater switching of suppliers than might be expected to occur in practice.
Number of potential suppliers	Adding more suppliers of the identical good as part of the experiments would result in more complexity in decision-making. The effect of having the number of potential suppliers beyond a given threshold is unlikely to be instructive in an experimental setting.	All subjects can choose from four suppliers. The project team determined that offering a choice of four allowed for enough variation of results without dramatically increasing the complexity of choices.

Many of the parameters described in Table 6 could form the basis of treatment options in a carefully designed experiment. As highlighted in Chapter 3.1 in the *Final Report*, the focus of the treatments chosen for the current experimental investigation was on

aspects under the direct control of the department in designing and implementing inspection rules. While there is no doubt that factors such as existing supplier relationships, supplier-switching costs and knowledge of inspection failure rates affect the practical operation of compliance-based inspection rules, these factors are largely outside the department's control[31] and can be taken as given by the department in devising its regulatory framework. That said, further investigations in terms of economic theory and experiments could aid the department's understanding of how these factors influence the scope of potential behavioural responses to the introduction of compliance-based inspection protocols.

As an initial simplifying assumption, the consignments offered by different suppliers are taken to be identical in all respects (for example, the amount and quality level) except for their landed cost and the likelihood of biosecurity risk material being present. While this represents a significant abstraction from what occurs in reality, it enables the experiments to identify the influence of biosecurity-related considerations on importer behaviour, holding other factors constant.

In practice, an importer may know very little about the approach rate for biosecurity risk material in consignments for particular supplier. This partly reflects the absence of a formal feedback mechanism provided by the department under the CBIS.[32] Such uncertainty could be captured by a *prior probability distribution*,[33] which could then be updated based on observed inspection outcomes. The process around how importers learn about a supplier's approach rate could be an important issue in the field if the main vehicle for improving biosecurity compliance is through selecting compliant suppliers.

However, from an experimental standpoint, the question of how importers form beliefs about the biosecurity standards of alternative suppliers and the way in which these beliefs are updated is distinct from how the design of the inspection protocols affects importer behaviour. To ensure the experimental results focus on evidence about the effects of different inspection protocols, we propose only investigating the situation where an importer "knows" a supplier's approach rate to be a specific value (for example, 10 per cent). This avoids the potential confounding effects and an additional source of variance in experimental results that learning dynamics could introduce.[34]

While assuming there are no costs involved in switching suppliers, the experiment becomes significantly simpler to explain to the subjects. However, stakeholder

---

31 How the department implements the regulatory framework could influence some of the factors listed in Table 6. For example, importers could learn the approach rates of alternative suppliers more quickly if the department provided them with tailored feedback on supplier-specific failure rates. This implementation strategy could also reduce the information search costs associated with switching suppliers, though transaction-related financial costs would still remain as a source of friction aiding supplier inertia.

32 In contrast, the legislative framework for the Imported Food Inspection Scheme (IFIS) allows for information on inspection failures to be more widely publicised under a "name and shame" approach. Information on testing failure rates by product are also publicly available through regular departmental publications.

33 A prior distribution in this sense would be the beliefs an importer held about a supplier before receiving consignments from them. This notion appeals to a theoretical construction in game theory of Bayesian updating of beliefs, as discussed in Appendix C of Rossiter et al. (2016).

34 A different form of experiment may be able to assess how importers "learn" about supplier failure rates and the implications this has for importer behaviour.

discussions as part of CEBRA Project 1304C suggested may importers perceived the costs associated with changing their suppliers and/or customs brokers to be considerable, and in some cases almost prohibitive. This results in a significant degree of inertia in the choices importers may make in practice, even if there might be significant cost benefits from changing to a supplier with a better track-record of compliance. Another implication of switching being costless is that no effective way of incorporating "default" options into the supplier choice context. In an environment where switching suppliers is costly, the use of default options can provide another way to examine the effect inertia has on impeding the type of "positive" behaviour change the department may seek to encourage.

In addition to the parameters described in Table 6, there are other parameters (Table 7) that influence importer decision-making whether they are vertically integrated or have a choice of suppliers. By assuming the regulator has perfect decision-making capacity regarding inspections, the last three parameters listed in Table 6 do not play an active role in our experimental design.

A consequence of restricting the choice set of suppliers to be common across all treatments is that the scope for introducing different incentives for compliance between importers rests on differences in the costs of inspection and treatment.[35] This implies that any testing of the "menu of regulatory contracts" approach advocated in CEBRA Project 1304C will rely on choosing different CSP rule parameters so that it is advantageous for one "type" of importer to choose one and another type to choose a different rule from a simple two-option menu.

Experimental validation for the menu of regulatory contracts approach is important to ensure that the incentive structures separating the different types of firms are sufficiently "sharp" to allow firms to appropriately choose which risk "bucket" they should be in, while ensuring that their actions are consistent with the biosecurity objective. Ideally, the inspection protocols could be structured so that importers and their suppliers should still be encouraged to improve their mitigation strategies relative to the case where all consignments are inspected. These issues are considered in more depth when considering the calibration of these rules.

[35] When broader notions of "costs" are considered, so they incorporate delay and reputational costs, these costs represent private information importers have about the parameters that affect their behaviour and are, in practice, difficult to observe.

Table 7: Parameters influencing importer choices when they have choice of supplier

Parameter	Comments	Use in the experiments
Resale price of goods in the domestic market	It is also implicitly assumed that any consignment imported is sold to the domestic market at this value.	Assumed to be the same for all consignments regardless of the supplier. This ensures that suppliers only differ in terms of the landed cost of goods and biosecurity risk characteristics.
Cost of a physical inspection	Includes direct charges for inspection as well as delay costs induced by needing to go through the inspection process. A higher cost of inspection provides a greater inducement for the importer to avoid inspection.	Two levels are used in the experiment: low (for treatment variation) and high (used as the reference level).
Cost to the importer of rectifying consignments with biosecurity hazards (treatment cost)	Includes direct costs associated with treatments to remove biosecurity risk material from a consignment. This could also capture indirect costs associated with affecting an importer's reputation as a "clean" importer and proxy for loss of future sales. Higher rectification costs are likely to result in steps being taken by importers that result in a lower likelihood of biosecurity risk material being present in consignments.	Two levels are used in the experiment: low (used as the reference level) and high (for treatment variation).
Probability of the inspector detecting hazards when present in a consignment subject to inspection	This parameter accounts for the potential for imperfect decision-making in the physical inspection process where an inspector fails to find biosecurity risk material that is present in a consignment. A lower probability of detection reduces the expected penalty faced by the importer of bringing in consignments with biosecurity risk material present. Acts to undermine the incentives for good compliance.	Taken to be one (i.e. "perfect" decision-making by the regulator) for the purposes of the experiment.
Probability of falsely detecting hazards when they are in fact not present	While under the control of the inspector, this reduces the expected difference in payoff between "clean" and "unclean" consignments. Similar to the other decision error of the inspector, this is expected to reduce the incentives for importers to comply with requirements.	Taken to be zero (i.e. "perfect" decision-making by the regulator) for the purposes of the experiment.
Cost of unnecessarily delaying consignments when hazards not present	This accounts for the situation where an initial inspection suggests biosecurity risk material is present, but subsequent investigation finds that risk material is not present. This adds to the costs of importers and may undermine their incentives to comply with biosecurity requirements.	Not applicable, given the event that causes this assumed to have zero probability (as per above).

3.3 Alternative experimental design to assess other aspects of biosecurity inspections

As noted throughout this report, this project has not provided an exhaustive examination of all possible appropriate dimensions of interest to a regulator in this environment. The following section briefly outlines some potential directions for additional assessments in an experimental setting that could be useful for policymakers in determining particular aspects of regulatory frameworks. These ideas are in addition to investigations assessing the role of alternative feedback systems for regulatory performance and on different ways to gain understanding of the rules.

3.3.1 Regulation in a social environment

The experiments conducted as part of this project assumed the regulator was able to commit to an inspection rule over all periods and stick to it regardless of the importer's actions. This extreme form of regulatory commitment is unrealistic, given the flow of new information the regulator will receive over time about changes in the risk profile of pathways over time, and unlikely to be optimal from a societal perspective.

An alternative experimental setup involving this type of social interaction would have some players taking on the role of importers and others being regulators. In this environment, the subjects playing the role of a regulator would make decisions around the inspection rules and would have the opportunity to respond to decisions made by importers. Such an experiment would allow a better understanding of the consequences for importer behaviour of allowing some discretion to the regulator, particularly when compared with the regulatory commitment framework used in this experiment. Furthermore, it could allow factors important for regulator decision-making, such as the trade-off in costs between biosecurity risk material leakage and undertaking an inspection, to be investigated in a systematic manner. The department could also learn about the application of mechanisms to tackle repeated non-compliance through appropriate penalty structures.

3.3.2 Accommodating differences between importers

Discussions with biosecurity system stakeholders and evidence from the department's administrative databases suggested that importers differ on a number of dimensions that were not investigated as part of this set of experiments. These include that importers:

- have different knowledge of or access to certain suppliers. A restricted choice set is an example of various information imperfections that are likely to exist in practice. The way in which regulatory schemes could help importers identify "good" supplier choices might then be very different from the way in which other issues are accommodated;
- may incur costs associated with switching suppliers. This means that an importer's "default" position can matter significantly for their subsequent choices and affect the likelihood of observing behaviour change in response to new rule settings. To overcome these switching costs, the department may need to think about implementing rules that have stronger incentives for compliance; and

- have different time horizons or volumes. An implication of the discussion in Chapter 3.5 regarding calibrating inspection rules is that protocols designed around high-volume imports may not be sufficient to encourage behaviour change among importers who bring goods into Australia less frequently.

Alternative experimental designs could enable the department to learn about the influence of these aspects of importer behaviour and potentially identify protocols that can assist the operation of rules under these differences in practice. For example, experimental treatments related to supplier-switching costs could be developed that allowed for different cost levels, which could allow behaviour to be compared with the situation without switching costs, and for different default suppliers, given that the status quo position matters if switching is costly. It is worth noting that this potential experimental setup could result in a substantial number of treatments, meaning an entire experiment may need to be devoted just to assess the influence of switching costs on the operation of compliance-based inspection protocols.

3.3.3 Understanding the transmission of regulatory incentives through the supply chain

The current experimental setup has characterised suppliers as fixed entities offering "take it or leave it" arrangements to domestic importers. However, stakeholder interviews undertaken as part of CEBRA Project 1304C indicated some importers have more collaborative relationships with their suppliers, with a considerable amount of communication between the two parties on biosecurity-related matters. Conceivably, introducing regulatory incentives for compliance on importers could in turn encourage "strategic" actions by suppliers to improve biosecurity risk mitigation practices at their end.

If developing a better understanding of the "transmission mechanism" for compliance-based regulation is of interest, then an experiment could be designed where both importers and suppliers form the participant groups. Importers would still be responsible for choosing their suppliers, while those playing suppliers could choose their cost structure and biosecurity compliance mechanisms as part of crafting offers to importers.

From the department's perspective, how the incentives that apply to importers "trickle down" the supply chain to affect the operation of other parties is a second-order issue. This is because the legal incidence for compliance with Australia's biosecurity requirements rests with the importer. Despite likely limited interest from biosecurity regulators in this perspective, there may be broader interest in this type of approach from regulators who need to operate by indirectly influencing third parties to instigate behaviour change and compliance with regulatory requirements.

3.4 Attitudes to risk and their influence on biosecurity choices

A key feature of the experimental economics approach involves the experimenter being able to manipulate subjects' payoff structures and assess their actual behaviour against specific theoretical predictions under different rules. However, as acknowledged by Holt and Laury (2014), calculating the "optimal" strategy under different treatments requires the experimenter to make assumptions about subjects' preferences. These preferences, including attitudes to risk, are beyond the

experimenter's direct control and their effects cannot be alleviated through processes such as random assignment to treatment.

Individual attitudes to risk form a key part of this experiment. In its most abstract construction, the main experimental task seeks to understand how people make choices in response to gambles with different payoffs and different probabilities attached to outcomes. In this situation, our interest is not about obtaining robust estimates of behavioural parameters controlling risk aversion *per se*; rather, it is to incorporate appropriate *controls* for the influence of individual risk preferences when subjects make supplier decisions. In this way, it becomes possible to obtain estimates of the treatment effects that are unaffected by differences in underlying attitudes to risk between subject groups.

3.4.1 Theoretical foundations of risk preferences

In the experimental setting, subjects face uncertainty choosing their supplier because it is not known *a priori* whether a particular consignment contains biosecurity risk material; instead, only the probability is known to the subject. For each good, the importer faces a *gamble* over a finite set of outcomes, namely whether the consignment is:

- not inspected;
- inspected and found not to contain biosecurity risk material; and
- inspected and found to contain biosecurity risk material,

as determined by the applicable inspection rule from the CSP family.

Each of these outcomes involve different payoffs for the subject. In linking the probabilities of the different outcomes and each player's preferences over those outcomes, the theoretical objective function for the importer can be represented by a *von Neumann-Morgenstern (vNM) utility function*.[36] As part of calibrating the theoretical framework underpinning the experiment, solving for the importer's optimal strategy in this context involves maximising the *expected utility* of the importer.[37] The form of a subject's vNM utility function will determine their preferences over the outcomes of the gamble.

In practice, there are sound reasons to think importers could be *risk averse*, preferring to receive the expected return of the gamble than face the risk inherent from the gamble itself. The likely distribution of risk preferences in the experimental subject pool is less clear; however, because different preferences for risk could result in substantially different "optimal" behaviours in the experimental setting, it is important to control for the influence risk appetite has on behaviour.

[36] A more formal discussion of decision-making under uncertainty and risk aversion using the expected utility theory paradigm can be found in standard microeconomic theory textbooks; for example, see Chapter 2.4 of Jehle and Reny (2001). See Harrison and Rutström (2008) for a discussion of characterising risk attitudes under alternatives to expected utility theory.

[37] While the expected utility theory framework is theoretically appealing, several experimental studies have found evidence of people not behaving as if they are expected utility maximisers; see, for example, Harrison and Rutström (2009).

3.4.2 Eliciting measures of risk attitudes in an experimental setting

A wide variety of experimental methodologies have been developed to elicit and assess individual risk attitudes in the laboratory and field.[38] However, determining which approach is most appropriate to use is a highly contested area in experimental economics.

Holt and Laury (2014) discusses three broad approaches to eliciting risk attitudes,[39] namely those based on:

- a single choice between alternative gambles (the *investment portfolio* approach; for example, Eckel and Grossman (2002, 2008));
- choices made according to a structured list of distinct binary choices between gambles with different risk characteristics (the *lottery choice menu*; for example, Holt and Laury (2002, 2005)); and
- the selling (or purchasing) price for a given gamble nominated by the subjects (the *pricing task* option; for example, Becker et al. (1964)).

These types of tasks are usually incentivised, so that subjects would be paid according to the outcome of risk elicitation task, with the available outcomes in part influenced by their choice. In addition to these more quantitatively oriented task, another approach that has gaining attention in the literature has been to ask a question around general attitudes to risk.[40]

The competing approaches to risk elicitation have their merits and deficiencies. Some aspects that need to be considered when selecting approaches to eliciting risk attitudes include:

- certain measurement approaches may have *noisier* or *less stable* responses than others, in that subjects may give different answers when asked the same question on several occasions. Hey et al. (2009) finds that responses to pairwise choice tasks closely related to the investment portfolio-type task are much less noisy than to pricing tasks, while Dulleck et al. (2013) finds only modest retesting consistency of decisions for examples of the lottery choice menu and pricing tasks.[41] In contrast, Lönnqvist et al. (2015) suggests that questionnaire-based measures may be more stable on retesting;
- potential *measurement biases* being invoked by certain risk-elicitation approaches. Among the pricing tasks, Hey et al. (2009) notes consistent under-bidding in willingness-to-pay measures and over-asking in willingness-to-accept measures for obtaining certainty-equivalent values. In contrast, they note that the more complex mechanism of Becker et al. (1964) appears to be neutral, but its complexity may induce participants to adopt simple heuristics that biases behaviour;

---

38 More comprehensive discussions of different risk attitude elicitation strategies can be found in Charness et al. (2013), Harrison and Rutström (2008) and Holt and Laury (2014).

39 The review article by Holt and Laury (2014) also provides a detailed discussion of the methodological issues, treatment effects and demographic patterns associated with risk preference measurement and assessment.

40 See Dohmen et al. (2011) for a comprehensive discussion of how this approach compares with incentivised tasks for risk elicitation.

41 Anderson and Mellor (2009) contend that that unobserved subject characteristics, such as comprehension or effort, could also influence measured risk preference stability.

- the *numerical aptitude* of the subject pool. Dave et al. (2010) find that for subjects with low-level mathematical skills that the simpler, yet coarser, Eckel and Grossman (2002) risk-elicitation approach is less noisy and seems to perform as well in a predictive sense relative to the more complex Holt and Laury (2002) approach. In contrast, they suggest that the finer-scale lottery choice menu procedure may be more appropriate if the subject pool has high-level mathematical skills;
- results from alternative approaches to measure risk attitudes may be, at best, only *weakly correlated*. This conclusion has been drawn across a wide range of risk-elicitation approaches[42] and may reflect the relatively low stability of measures on retesting (Dulleck et al., 2013). The weak correlation of alternative measures can lead to substantially different characterisations of the relevant experimental population's risk attitudes (see, for instance, Hey et al. (2009)) and could also mean some risk attitude measures bear little correlation with actual risk-taking behaviour in related experimental tasks (see, for instance, Lönnqvist et al. (2015)); and
- the ability for the measures to be *generalised to other risk-taking contexts*. According to Dohmen et al. (2011), a survey-based global assessment of willingness to take risks appears to generate a "useful all-round measure" of attitudes to risk. More narrow assessments, where a question focuses more on lottery or investment choices pertinent to the financial decision-making domain, are likely to be highly predictive of behaviour for that specific context, but may be less useful in identifying risk attitudes in other domains.

Given these competing considerations, Charness et al. (2013) suggests that the approach, or approaches, appropriate in a specific context will largely depend on the research question of interest and the characteristics of the population from which the experimental subjects are drawn.

For this experiment, the project team chose a variant of the first type of mechanism based on the choice task described in Eckel and Grossman (2002, 2008). This task is relatively simple to administer[43] and provides good information about variation in the degree of risk aversion across subjects. In this setting, we use the no-loss formulation of the risk-elicitation task to avoid complications arising from potentially different responses to losses versus gains.[44] The actual parameters used in the risk task for our experiment are given in Table 8, where the dollar values represent the actual monetary rewards received by subjects, dependent on the experimental outcome.

[42] See, for example, Anderson and Mellor (2009), Dulleck et al. (2013), Hey et al. (2009) and Lönnqvist et al. (2015) and the references therein.

[43] In considering the subject pool, the project team acknowledged that students would be drawn from a wide range of disciplines, not necessarily with a strong numerical focus. As such, the Eckel and Grossman (2002) task would be more appropriate than the Holt and Laury (2002) task in this context.

[44] See Tversky and Kahneman (1991) for a discussion of loss aversion and reference-dependent consumer choice.

Table 8: Gamble choices, expected payoffs, risk and underlying risk parameter ranges

Gamble choice	Event	Probability (%)	Payoff ($)	Expected payoff ($)	Risk˜ ($)	CRRA parameter ranges*
1	A	50	4.50	4.00	0.50	r > 2.33
	B	50	3.50			
2	A	50	6	4.50	1.50	0.67 < r < 2.33
	B	50	3			
3	A	50	8	5.00	3.00	0.38 < r < 0.67
	B	50	2			
4	A	50	10	5.50	4.50	0.20 < r < 0.38
	B	50	1			
5	A	50	12	6.00	6.00	r < 0.20
	B	50	0			

Notes:
˜ The standard deviation of payoff probability distribution.
*This is calculated as the range of *r* in the function:

$$U(x) = \frac{x^{1-r}}{1-r}$$

for which the subject chooses each gamble (excluding the show-up fee) assuming constant relative risk aversion (CRRA) utility. Under this parameterisation of the CRRA utility function, a subject displays risk aversion if r>0, risk neutrality if r = 0 and risk-loving behaviour if r < 0.

The risk-elicitation task is the first task the subjects perform in the sequence of tasks in this experiment. By placing this before the supplier-choice experiment, we avoid the potential of conditioning responses based on their performance in the experiment. This should allow for a more "pure" measure of risk preferences and reduce potential ordering effects. The experimental subjects only find out the results from this task after completing the main biosecurity supplier-choice task.

In addition to the incentivised risk-elicitation task, we also ask participants to rank their general attitudes to taking risks in the post-experimental questionnaire, in line with the survey-based risk measures espoused by Dohmen et al. (2011), among others. Lönnqvist et al. (2015) notes that these survey measures are useful as predictors of behaviour, particularly where relative assessments of risk attitudes are more the focus. However, since this measure is not incentivised, some would argue it is debatable "whether the elicited risk preferences reflect an individual's true attitudes toward risk, particularly in the domain of financial decision-making" (Charness et al., 2013, 45) – the relevant domain for the biosecurity regulation experimental task. Based on the previous discussion, it would not be surprising if the results of the survey-based and incentivised measures to elicit risk attitudes employed in this experimental study were not strongly correlated, or if one or both measures were of limited use as a statistical control in analysing supplier choices in the main experimental task.

3.5 Developing benchmarks for predicting behaviour in the biosecurity experiments

The design of experimental treatments is a non-trivial exercise. Earlier sections in this chapter have highlighted how many sources of influence on importer decision-making are being held constant across treatments; however, there remain many features that can still be explored.

This section describes the approach used to develop benchmarks for predicted behaviour in the experiments and how the various parameters in the theoretical model were calibrated for implementation in the experiment. We also describe the 18 treatments used as part of this experiment, discussing how these treatments can then be compared to allow inferences about the implications for regulatory mechanisms and the way in which biosecurity regulations can be implemented.

To identify theoretical predictions for the "baseline" model and for various treatment effects, members of the project team developed a simulation model[45] to determine the "optimal" strategies for subjects under the assumption they were expected-utility maximisers.[46] As noted in Rossiter and Hester (2017), a feature of the CSP family of rules is that they are path-dependent, meaning analytical solutions to the optimal strategies to be pursued by subjects are difficult to determine. Similar to the strategy adopted in that work, we assume the optimal strategy takes the form of a "switching" algorithm, where the importer can choose different suppliers when in the census and monitoring phases of the CSP rule.

When selecting parameters for experimentally testing some theoretical model predictions, an ideal situation would be where:

- the payoff function, which maps the strategies subjects could employ in making choices to the monetary rewards they receive, has a sufficiently steep gradient to admit relatively clear theoretical predictions;
- there are marked differences between the theoretical predictions associated with different treatments; and
- difficult-to-measure underlying influences on behaviour, such as individual risk preferences, have little or no bearing on the predicted solution.

However, it became clear through trialling various model parameterisations that the CSP rules did not provide strong incentives for compliance under cost parameters that were broadly consistent with intelligence from stakeholder interviews in CEBRA Project 1304C. In particular, it was difficult to find cases where the differences in payoffs between the worst and best supplier-choice strategies were marked, let alone the optimal strategy providing a significantly larger payoff than the next best strategy.

Greater separation of payoffs was possible under circumstances where inspection and/or treatment costs were very high. However, as noted in Chapter 3 of the *Final Report*, these situations may occur in relatively few real-world cases, given the

45 The MATLAB files used to generate the output discussed in this section are described and documented in Appendix B of this report.

46 These computations also assume the subjects are fully informed of the CSP rule parameters and they have the option to switch suppliers after each consignment. For simplicity of exposition, we present the average payoff estimates for a risk-neutral experimental subject. For risk-averse experimental subjects, their optimal strategy tends to favour selecting suppliers with lower approach rates.

restrictions placed on these cost structures by other Australian Government policy considerations and international agreements. For completeness, we illustrate this type of case later in the section.

3.5.1 Establishing parameter values for biosecurity inspection task without rule choice

The final parameterisations arrived at by the project team attempted to ensure the payoff functions covered a reasonably wide range, so as to provide an appropriate monetary inducement for students for making "better" choices in the experiment. This was done within the confines of having to try to keep a large number of parameters constant across treatments, to ensure the number of treatments remained workable. It also considered parameter values that would be simpler for subjects to easily comprehend (i.e. rounded whole numbers) and allow them make simple calculations if they opted to do so.

The final set of parameters used in the experiments are summarised below in Table 9 for the supplier characteristics and costs common to all treatments.

Table 9: Supplier options in the biosecurity inspection experiment

Supplier option	**A**	**B**	**C**	**D**
Transportation and purchase costs per good (in monetary units)	**3**	**4**	**6**	**8**
Probability that a good in a shipment contains biosecurity risk material***	**50%**	**30%**	**10%**	**2%**

*** Note that a probability of, for example, 50% does not automatically imply that 5 out of the 10 goods in a shipment contain biosecurity risk material but that there is a 50% probability that each single good contains biosecurity risk material. Thus, it is possible that the number of goods in a shipment containing biosecurity risk material is less than, equal to, or greater than 5.

According to the assumptions described in earlier, the suppliers only differed with respect to their landed costs of goods and the approach rate of biosecurity risk material. There was a negative relationship between the two dimensions.[47] Suppliers with a lower probability of biosecurity risk material had higher shipment costs.

Table 10 lists the other model parameters determined as part of the calibration exercise.

[47] Without this restriction, some supplier choices could be "dominated" by others and would not be chosen by a "rational" agent, regardless of their risk preferences. This would result in the experiment becoming a test of for rationality, rather than considering how to use incentives and behaviour-based interventions designed to encourage behaviours consistent with a regulatory objective.

Table 10: Calibrated importer cost parameters and regulator rule parameters

Importer cost parameters	Value (monetary units)	Regulator rule parameters	Value
Resale price of goods	20		
Cost of physical inspection		**Clearance number**	
Low (variation)	2	Baseline	10
High (baseline)	4	Alternative choice	5
Treatment cost		**Monitoring fraction**	
Low (baseline)	6	Baseline	0.2
High (variation)	12	Alternative choice	0.3

Figure 9 provides an illustration of the difficulties associated with calibrating the CSP rules. This shows the average value of the payoff function based on 1 000 simulations for different supplier-choice strategies[48] under the "baseline" treatment (henceforth referred to as treatment C1), which uses the CSP-1 rule (under full information) and the parameters from Table 10 referenced as "baseline" parameter values. These simulations allow for 500 supplier choices, each for one consignment, with the payoffs not subject to discounting.

The average payoff function value across different supplier-choice strategies is relatively flat. In this case, the "optimal" strategy (choosing supplier D when in census mode and supplier C when in monitoring mode) has an average payoff only 20 per cent higher than the "worst" strategy (in which the importer would choose supplier A in both modes of the algorithm). There are three other supplier-choice strategies, namely (C,C), (D,B) and (C,B), which are all within 1.5 per cent of the average payoff of the "optimal" strategy. It is noteworthy that the supplier-choice strategies (C,B) and (D,B) would imply significantly higher approach rates for biosecurity risk material relative to the optimal strategy (D,C). This implies that the importer gains only a small incremental return by pursuing the optimal supplier-choice strategy consistent with a lower average approach rate over other strategies that result in a greater likelihood of biosecurity risk material leaking into the environment.

[48] As described in Chapter 4.1 of the *Final Report*, the supplier strategies are presented in the form of the pair *(x, y)*, where *x* is the "best" supplier choice under census (100 per cent) mode of the relevant CSP rule and y is the choice under monitoring mode. For example, the predicted optimal strategy under treatment C1 (shown in Figure 9) is to choose supplier D when subject to mandatory inspection and then choose supplier C when subject to the 20 per cent inspection rate in monitoring mode.

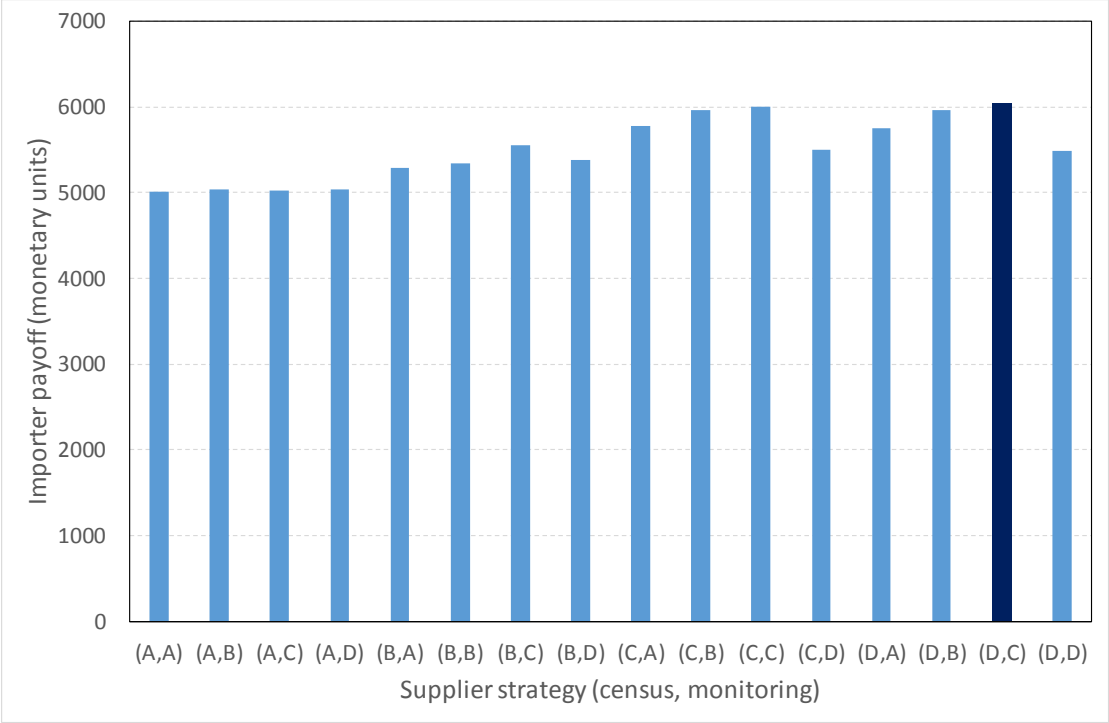


Figure 9: Calibrated expected payoff values for the "baseline" CSP-1 treatment (C1)

Note: The "optimal" strategy (D,C) for the importer has been recoloured for emphasis.

What Figure 9 does not show is the substantial variability in the realised payoffs possible under different supplier-choice strategies. Figure 10 shows histograms of the realised payoffs for the four strategies – (C,B), (C,C), (D,B) and (D,C) – which have the highest average payoffs shown in Figure 9. For emphasis, the strategy with the highest average payoff, (D,C), is highlighted with a deeper tone of blue.

The realised payoff values cover a wide range, from around 5 200 to 6 800 monetary units,[49] with considerably more payoff variability for the strategies choosing supplier B in monitoring mode (top-left and bottom-left panels). The variability of payoff realisations reduces considerably when choosing suppliers with lower approach rates, particularly in monitoring mode, as suggested in the top-right and bottom-right panels of Figure 10. For instance, the standard deviation of payoffs for strategy (D,C), at 169 monetary units, is almost one-third lower than the standard deviation for strategy (D,B), at 245 monetary units. Notably, the standard deviations of payoff distributions are large compared to the differences between the expected payoffs, again suggesting that it may be difficult to observe clear behaviours in an experimental context.

[49] In the experiment, payoffs received by subjects in treatment C1 (across all strategies) ranged from 5 276 to 6 262 monetary units, with a mean payoff of just under 5 800 monetary units.

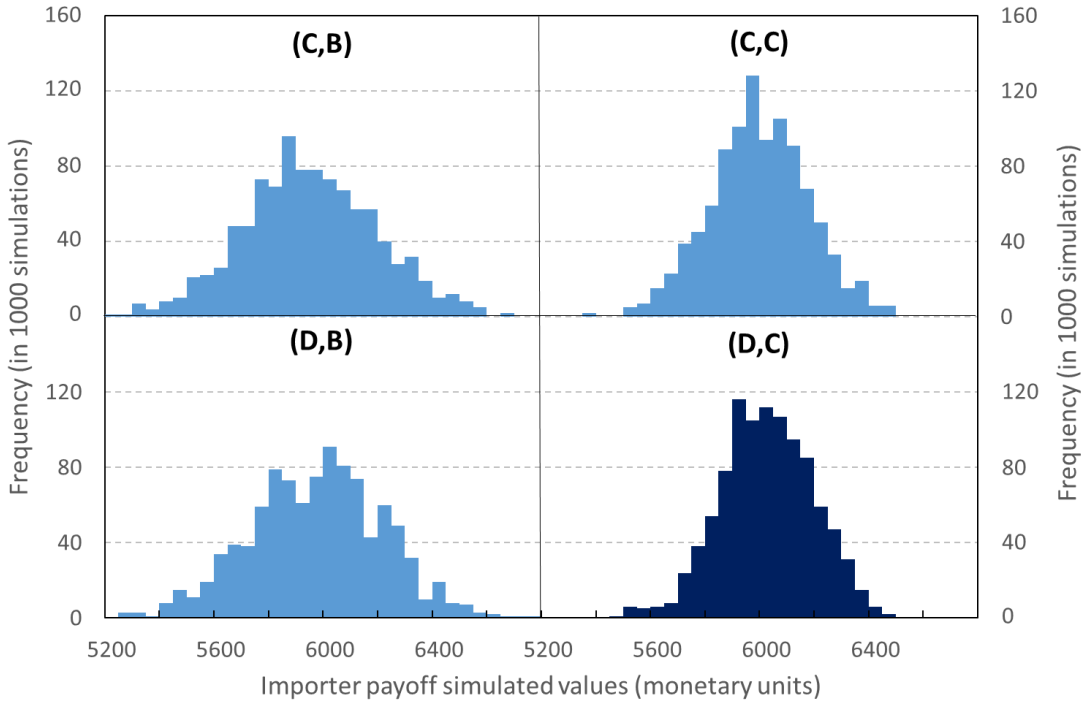


Figure 10: Histograms of payoff distributions for different supplier-choice strategies under treatment C1

3.5.2 Parameterising a second menu option for the choice environment

The difficulties in separating expected payoffs carried over into developing the menu of regulatory contracts based on rule parameter changes alone, as illustrated in Figure 11. This chart compares the average payoff functions for the two alternative rules faced by importers whose inspection and treatment costs were at the "baseline" levels. Note that we have chosen to rescale the chart (starting at 5000 monetary units) to highlight the small differences in payoffs between these two rules. In the case illustrated in Figure 11, it is in the interests of the subject to choose the CSP-1 rule with clearance number 5 and monitoring fraction 0.3, as it gives a larger expected payoff under the supplier-choice (C,B) than does the baseline rule pursuing strategy (D,C). As with treatment C1, treatment C1-5.03 has a number of strategies, notably (C,A), (C,C), (D,A), (D, B) and (D, C), that offer a payoff within 5 per cent of the average payoff under the "optimal" strategy. Without conducting this type of simulation exercise, it would be difficult for participants to establish which strategy would best suit their situation.

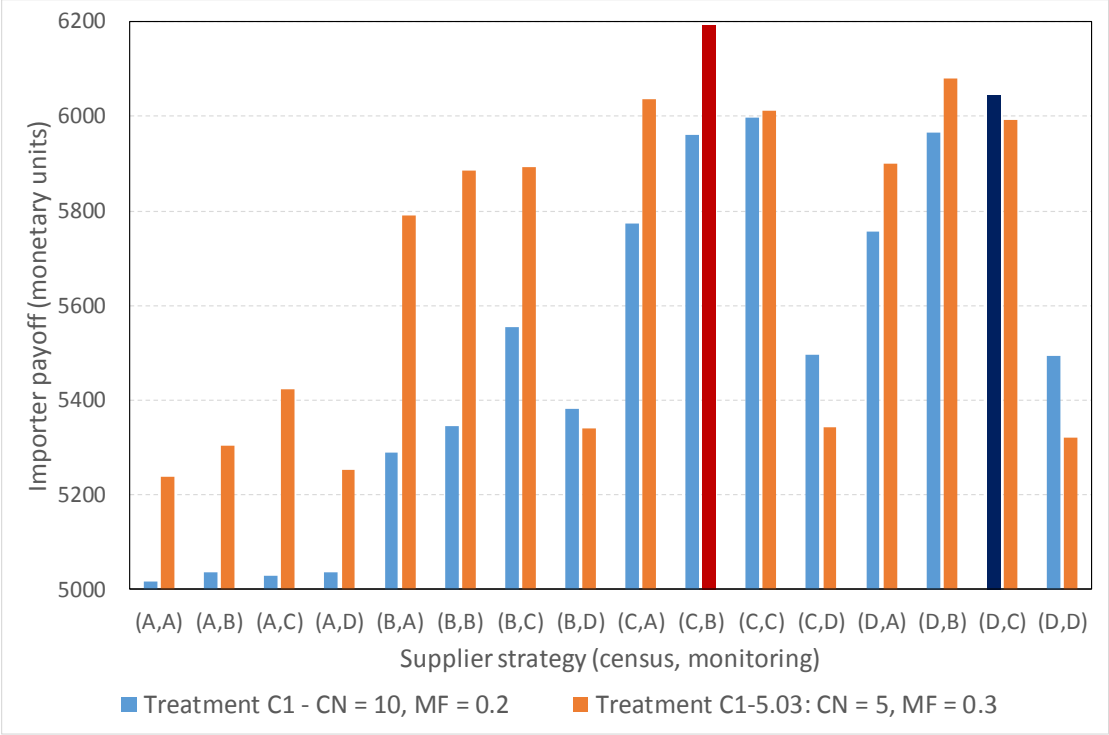


Figure 11: Calibrated expected payoff values for the baseline (C1) and alternative choice (C1-5.03) CSP-1 rules

Note: "Optimal" strategies for both schemes have been recoloured for emphasis.

Given the difficulties faced in calibrating rule options for a menu of regulatory contracts, it is unlikely that menus developed on the basis of offering rules with alternative CSP rule parameters alone will deliver benefits to the department. Instead, the project team acknowledges menus need to be carefully designed, taking into account the considerations noted in the predecessor project, CEBRA Project 1304C.[50] With this in mind, regulatory options are likely to be more successful if eligibility to "lighter touch" border inspection arrangements is based on a system stakeholder demonstrating biosecurity risk management practices known to reduce the approach rate of biosecurity risk material. Importantly, this adherence to risk management practices must be able to be independently verified through mechanisms such as internationally accredited audit programs. In adopting these eligibility "thresholds" for options with reduced intervention at the border, the department is more likely to encourage actions in the supply chain consistent with its aims of reducing the likelihood of pest and disease incursions and maintaining Australia's high biosecurity status.

3.5.3 Implications of alternative cost structures on importer behaviour: an illustration

For the purposes of illustration, we also demonstrate the effect of introducing a higher treatment cost (20 monetary units) on the baseline treatment scenario illustrated in Figure 9. This cost structure implies participants would lose money if one of their

[50] See, for instance, Chapter 4.3 of Rossiter et al. (2016). For a more detailed discussion of technical issues associated with designing menu of contracts, see Laffont and Tirole (1993) or the more policy-oriented Sappington (1994).

imported "goods" was inspected and contained biosecurity risk material for any supplier they chose. In our experiment, we deliberately sought to avoid this type of scenario in case it induced behaviour consistent with loss aversion among participants.

Figure 12 shows the payoffs for the available strategies in census and monitoring modes of the CSP-1 algorithm when the treatment cost is 20 monetary units. Relative to Figure 9, the range of payoffs is considerably larger; specifically, strategies associated with choosing supplier A or B in census mode deliver much lower payoffs when treatment costs are very high. The optimal strategy in this scenario is again (D,C), though strategies (C,C) and (D,B) deliver expected payoffs within 3.5 per cent of that strategy's payoff. This highlights the challenges of obtaining sharp theoretical predictions out of this type of interaction, even with "extreme" cost parameterisations.

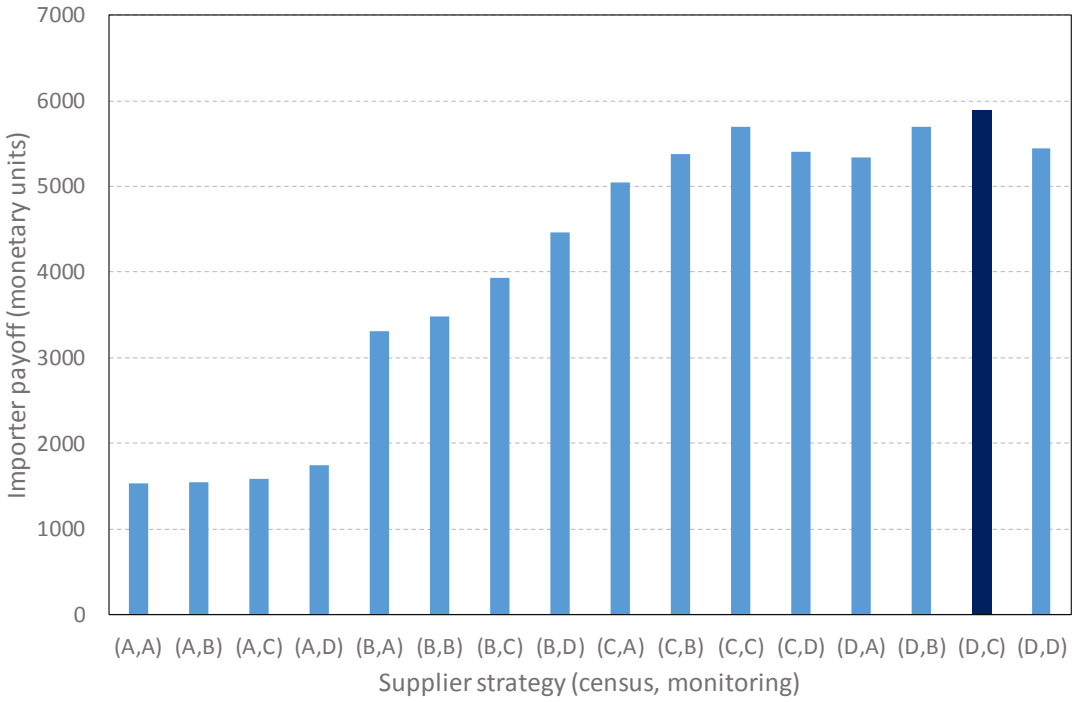


Figure 12: Expected payoff values for a very high treatment cost (20 monetary units) scenario

This "very high treatment cost" scenario is instructive from another perspective, as it reinforces arguments made in Rossiter et al. (2016) about the potential for behaviour changes to be observed in moving from mandatory inspections to compliance-based inspection protocols. Under mandatory inspections with very high treatment costs, it can be easily verified that an importer's "optimal" response would be to choose supplier C, with supplier D a close second choice. Because importers already have a strong incentive to avoid failing inspection under the mandatory inspection regime, changing protocols so inspections follow a CSP-1 algorithm with $CN = 10$ and $MF = 0.2$ would not elicit much behaviour change in this scenario.

In contrast, the parameters associated with the baseline scenario described in Table 10 yield a predicted choice of supplier B under a mandatory inspection regime, with supplier A providing the second highest expected payoff. Introducing a compliance-based inspection protocol in this context could induce noticeable behaviour change by importers. This is because the potential to avoid inspections and their associated costs encourages switching to suppliers with lower approach rates,

with (D,C) is the optimal strategy. Thus, it is possible to develop scenarios where compliance-based protocols could provide sufficient rewards to encourage behaviour change consistent with regulatory objectives. Whether this is "an improvement" from the regulator's perspective will ultimately depend on their relative costs of undertaking inspections and leakage of biosecurity risk material.

4. Experimental Data Overview

In this chapter, we preview the data collected during the experimental sessions conducted as part of CEBRA Project 1404C. In total, 275 individuals took part in our experiments in 12 experimental sessions conducted over September and October 2015 in the Monash University Laboratory for Experimental Economics (MonLEE) in Clayton, Melbourne. The number of subjects in the experimental sessions varied between 19 and 24 individuals.

We briefly describe the characteristics of the subject pool, before turning to examine the response from the first three tasks in the experiment in turn.

4.1 Characteristics of the subject pool

Participants in the experiments were students from different disciplines at Monash University. For these experiments, there tended to be a particular concentration of students whose major was either accounting or engineering. The table also highlights the over-representation of students from Asia – a feature we explore further in Chapter 6. Other interesting characteristics of note for the subject pool are described below in Table 11.

Table 11: Characteristics of experimental subjects participating in the biosecurity experiments

Characteristic	Attribute	Number of participants	Share of experimental subjects (%)[~]
Gender	Female	124	45.1
	Male	151	54.9
Nationality	Australian	61	22.2
	Asian[^]	193	70.2
	Other[#]	21	7.6
High school attended	Public	139	50.6
	Private	84	30.6
	Catholic	17	6.2
	International	35	12.7
Employment	Held a part-time job	75	27.3
	Average hours worked per week in part-time job	16 hours	*

Notes:
[~] Percentage totals may not add to 100 per cent due to rounding.
[^] Within this category, the main countries of origin were China and Malaysia.
[#] This category includes students from North and South America, Africa, Oceania and New Zealand.

4.2 Risk preferences

Before the participants started with the main experiment (the second task on supplier choices), they took part in a risk experiment. In this experiment, they had to select the gamble they most preferred out of a list of five, as described in Chapter 3.4.2.

Table 12 summarises the participants' choices in this task. We find considerable variation in participants' choices and that all five gambles were frequently chosen. However, the majority chose either:

- gamble 1, which had the lowest risk ($0.50) and lowest expected payoff ($4); or
- gamble 5, which had the highest risk ($6) and highest expected payoff ($6).

Table 12: Gamble choices in the abstract risk task

Gamble	Event	Probability (%)	Payoff ($)	Number of subjects choosing this gamble	Share of subjects (%)
1	A	50	4.50	75	27.3
	B	50	3.50		
2	A	50	6	44	16.0
	B	50	3		
3	A	50	8	46	16.7
	B	50	2		
4	A	50	10	28	10.2
	B	50	1		
5	A	50	12	82	29.8
	B	50	0		

In addition, we measured risk preferences in the post-experimental survey using a question asking participants about their general willingness to take risks on a scale from 1 (not at all willing to take risks) to 11 (very willing to take risks). Figure 13 summarises the survey responses, with most respondents choosing medium to high levels of willingness to take on risk – that is, responses between 4 and 9.

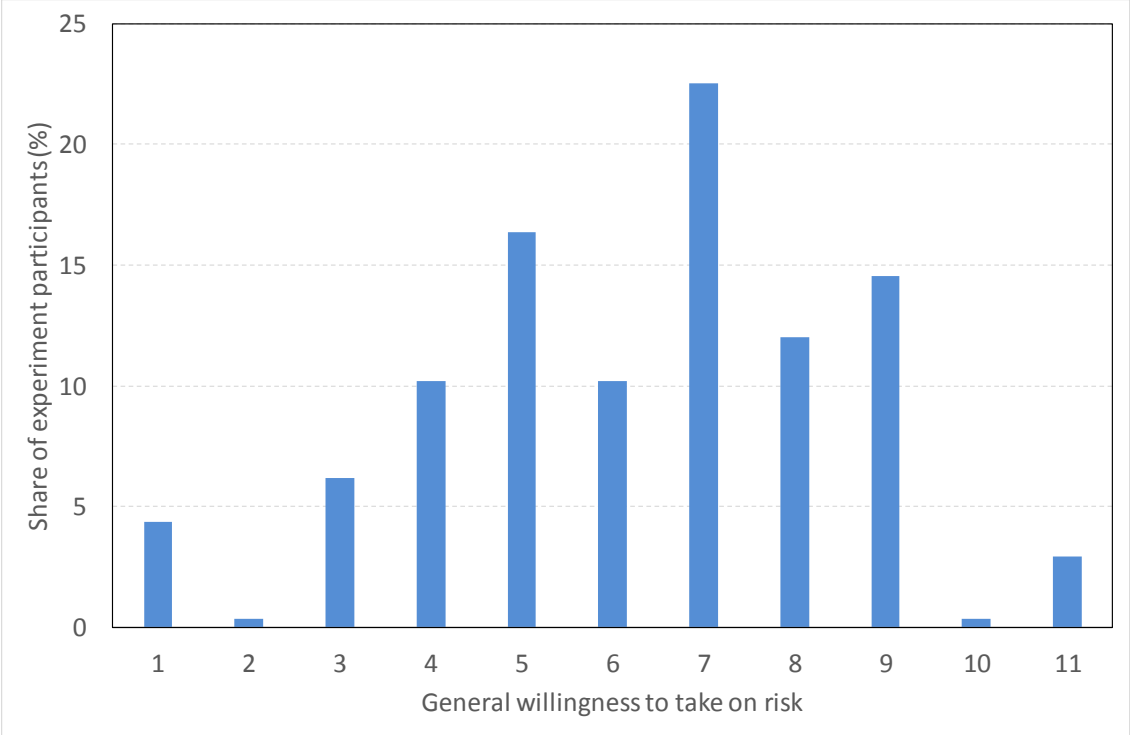


Figure 13: Self-reported willingness to take on risk from post-experiment questionnaire

Notes: The 11-point scale goes from not at all willing to take risks (1) to very willing to take risks (11).

While the risk experiment identifies a specific willingness to take risks in a financial gambling situation and the survey measure a more general willingness to take risks, the two ordinal measures are positively correlated with a Spearman rank correlation coefficient of 0.27.[51] This positive association is confirmed in Figure 14, which show the cumulative distribution functions of post-experiment questionnaire responses on the risk question by the gamble chosen in the abstract risk-elicitation task. For the most part, the cumulative distribution functions of the questionnaire responses for those who chose the gambles with higher levels of risk (4 or 5) tend to be below and to the right of those for the less risky gambles (1 or 2). This means fewer subjects who chose higher-risk gambles selected questionnaire responses indicating less willingness to take risks in general, while more chose responses that suggested a greater willingness to take risks. Somewhat surprisingly, the results for subjects who chose the medium-risk gamble (3) seem to contradict the general pattern, as the cumulative distribution function points to them being less willing to take risks in general than those who chose gamble 2 in the first experimental task.

The positive, but relatively small, correlation between these measures is broadly consistent with the results of Hey et al. (2009), which showed pairwise correlations to compare the outcomes of four types of incentivised risk-elicitation tasks. The results here, however, are more favourable than those found by Lönnqvist et al. (2015), where the survey-based and lottery choice menu measures of risk attitudes displayed little correlation.

[51] The p-value associated with a test of the hypothesis of no correlation against the alternative of the measures being positively correlated is less than 0.001.

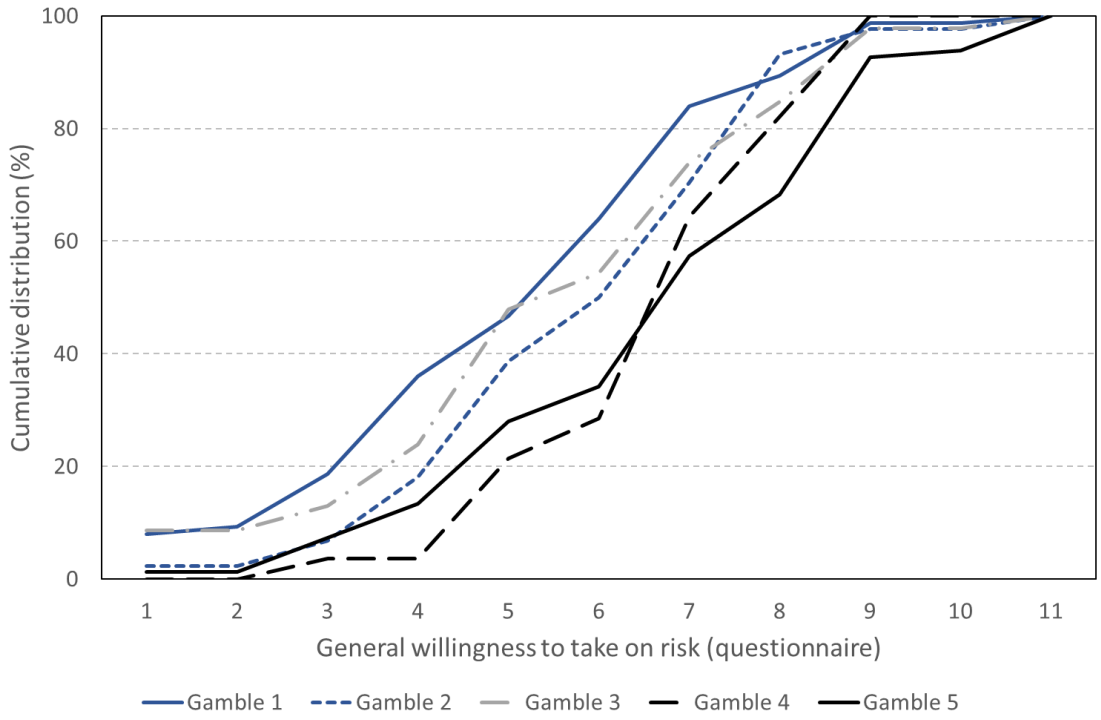


Figure 14: Cumulative distribution of self-reported willingness to take on risk from post-experiment questionnaire by gamble choice in abstract risk task

4.3 Supplier choices in the main experimental task

In the main experiment, participants had to select one out of four suppliers for 50 periods. The characteristics of suppliers were described in Table 9 in Chapter 3, with the suppliers differing on two important dimensions, namely:

- the costs per shipment; and
- the probability of each good containing biosecurity risk material.

Table 13 summarises the supplier characteristics and shows the participants' choices in all treatments according to different periods in the experiment.[52]

Table 13: Subject supplier choices over different time periods in the biosecurity inspection game

Supplier	A	B	C	D
Costs per shipment (monetary units)	30	40	60	80
Probability that a good contains biosecurity risk material (%)	50	30	10	2
Choice in all periods (%)	14.3	30.0	35.1	20.7
Choice in first period (%)	5.8	20.0	38.9	35.3
Choice in first 10 periods (%)	13.1	28.6	32.4	25.9
Choice in last 10 periods (%)	15.4	31.8	36.8	16.0
Choice in last period (%)	33.8	30.9	26.9	8.4

[52] In this experiment, we observe a total of 13 750 individual choices, comprising 50 choices per participant over 275 participants.

Over all treatments and in all periods, the most selected supplier was C, while supplier A was the least preferred. However, we also find that supplier choices shifted over periods towards suppliers with higher probabilities of biosecurity risk material. For example, in the first period only around one-quarter of participants chose suppliers A or B. In the last ten periods of the game, the share choosing supplier A or B rose to almost half, with the corresponding percentage rising to almost two-thirds in the last period. This type of behaviour, where participants make "riskier" choices towards the end of the game, is consistent with theoretical predictions for finite-horizon games. As outlined in Chapter 2, this reflects the expected cumulative reward for compliance decreasing as the game approaches its final periods, leading to decisions being based more on short-term reward structures.

In Chapters 5 and 6, this report investigates how supplier choices differed based on the experimental treatments and individual characteristics. As part of this analysis, we also control for period (time) effects.

4.4 Post-experiment questionnaire attitudinal responses

In the post-experimental survey, we collected additional information on participants' characteristics. Table 14 provides an overview of the questions and responses related to environmental concerns, political attitudes, and understanding of the experiment.

We find that participants hold relatively high concerns for the environment; for example, they tend to find it more important to protect the environment than to improve the economy. The individual responses to the three questions related to environmental concerns are also highly correlated, suggesting that environmental concerns are a relatively stable individual characteristic.

Two questions capture political attitudes (government interventions and liberal-conservative attitudes). We observe that, on average, individuals who participated in this experiment were somewhat positive inclined towards government interventions to solve environmental problems. Most participants consider themselves to be either moderately conservative or moderately liberal, with more than half of the experimental subjects choosing the middle three values (5, 6 and 7) on an 11-point scale for the question on their political attitudes.

Chapter 6 considers the effect of these characteristics on supplier choices in more detail.

4.5 Measures of rule understanding

4.5.1 Perceived versus demonstrated rule comprehension

As noted in Chapter 2, we used two measures for how well the experimental subjects understood the inspection rules – a self-reported measure from the post-experimental questionnaire and one based on results from the paper-based incentivised task conducted after the questionnaire.

Table 14: Post-experiment questionnaire aggregated responses

Question	Scale	Mean	Standard deviation	Additional information
How concerned are you about the risk of introducing exotic pests and diseases to Australia?	5-point scale, from "not at all concerned" (1) to "extremely concerned" (5)	3.40	0.94	Responses to these three questions are strongly correlated ($p<0.001$ on a rank-correlation test)
How concerned are you about the extinction of endangered plants and animals?	5-point scale, from "not at all concerned" (1) to "extremely concerned" (5)	3.76	0.99	
How important is protecting the environment compared with improving the economy?	7-point scale, from "much less important" (1) to "much more important" (7)	5.51	1.37	
When governments get involved in trying to solve environmental problems, how often do you think they make things better?	5-point scale, from "never" (1) to "always" (5)	3.09	0.86	
We hear a lot of talk these days about 'liberals' and 'conservatives'. Here is an 11-point scale on which people's political views are arranged from extremely liberal to extremely conservative. Where would you place yourself on this scale?	11-point scale, from "extremely liberal" (1) to "extremely conservative" (11)	5.85	2.20	More than half the subjects (55.4 per cent) chose the middle three options (5, 6 or 7)
How well did you understand the mechanism that the quarantine service used to inspect goods?	11-point scale, from "I do not understand at all" (1) to "completely understand" (11)	7.45	3.14	5.8 per cent do not understand at all; 22.2 per cent completely understand

Responses in the post-experiment questionnaire on rule understanding in the last row of Table 14 indicate that few participants (16 out of 275, or 5.7 per cent) reported not understanding the mechanism that the quarantine service used to inspect goods at all. In contrast, almost one-quarter of participants (61 out of 275, or 22.2 per cent) reported to completely understand the mechanism.[53] The average self-reported score was quite high, suggesting that participants felt reasonably confident in understanding the inspection rules they were subject to in the main experimental task.

In the paper-based task, 98 of the 275 participants (35.6 per cent) correctly answered both scenarios according to the rule they followed, with a further 16 participants (5.8 per cent) answering one of the two scenarios correctly. This meant more than half of the participants failed to correctly answer either scenario placed before them, casting significant doubt as to whether participants truly understood how the rules worked. It appears that subjects, in general, may have been overconfident in perceiving they understand the rules to a greater extent than they were able to demonstrate in an independent measurement of rule comprehension. This is not surprising, but highlights a potential challenge for biosecurity regulators to be able to clearly explain compliance-based inspection rules to stakeholders.

Table 15 compares the post-experiment questionnaire ratings and the paper-based task performance of subjects who were subject to CSP-1 or CSP-3 rules in the main experimental task. The results in the table show clear evidence of over-confidence by subjects, as only slightly less than half (41 out of 84) of the respondents who scored their understanding of the rules as 10 or 11 on the 11-point scale answered both scenarios correctly. There is some evidence of a positive association between the two indicators of rule understanding, with a Spearman rank correlation coefficient between the two measures of 0.19. Despite the positive association between the self-reported and objective measures of rule understanding,[54] these measures could suggest very different patterns when assessing how they influence supplier choices in more sophisticated panel regression models.

[53] There appears to be little evidence that the self-reported score patterns differed markedly by treatment. The p-value for Fisher's exact test of independence between treatment and self-reported understanding score was around 0.15, suggesting there was at most modest evidence of differences between the 18 treatments as to how subjects perceived to understand the inspection rules.

[54] The p-value associated with a test of the hypothesis of no correlation against the alternative of the measures being positively correlated is around 0.001.

Table 15: Post-experiment questionnaire responses versus paper-based task performance

Post-experiment questionnaire rating	Number of participants	Number of scenarios answered correctly		
		0	1	2
1	15	9	0	6
2	3	2	0	1
3	28	21	1	6
4	7	5	0	2
5	23	17	0	6
6	15	13	0	2
7	32	16	4	12
8	7	1	0	6
9	50	34	4	12
10	26	11	3	12
11	58	26	3	29

Notes: Excludes the 11 participants in treatments M and R (those with mandatory and randomised inspections, respectively) whose paper-based task and post-experiment questionnaire ratings related to different rules.

Table 16 highlights that the performance on the paper-based task differed quite markedly by treatment. The choice treatment with high treatment costs (treatment Choice12) was the only treatment where more than half the subjects answered both scenarios correctly, while other treatments (notably treatments C1-G and C1-5.03.12) had fewer than one-quarter of participants answering either scenario correctly. Surprisingly, participants in the CSP-3 algorithm treatments (treatments C3, C3-I and C3-I2) performed *slightly* better overall than those subject to the more straightforward CSP-1 algorithm on this task.[55] This was mainly because several participants in the CSP-3 treatments answered one of the two scenarios correctly.

[55] Note that Scenario 1, shows in Appendix A.4, was particularly designed to assess whether participants properly understood the difference between the CSP-1 and CSP-3 algorithm penalty mechanisms. For Scenario 2, the inspection process would be identical under the CSP-1 and CSP-3 algorithms.

Table 16: Paper-based task performance versus biosecurity experiment treatment

Treatment identifier	Number of participants	Number of scenarios answered correctly			Percentage answered correctly	
		0	1	2	At least one scenario	Both scenarios
M	5	3	1	1	40.0	20.0
R	6	3	0	3	50.0	50.0
C1-I	12	8	0	4	33.3	33.3
C1	21	11	2	8	47.6	38.1
C3-I	18	8	4	6	55.6	33.3
C3-I2	16	7	3	6	56.3	37.5
C3	23	12	3	8	47.8	34.8
C1-IL	17	9	1	7	47.1	41.2
C1-L	17	12	0	5	29.4	29.4
C1-IG	18	9	1	8	50.0	44.4
C1-G	18	14	0	4	22.2	22.2
C1-2.6	12	8	0	4	33.3	33.3
C1-2.12	12	6	0	6	50.0	50.0
C1-4.12	9	6	0	3	33.3	33.3
C1-5.03	17	12	0	5	29.4	29.4
C1-5.03.12	18	14	0	4	22.2	22.2
Choice6	18	12	0	6	33.3	33.3
Choice12	18	7	1	10	61.1	55.6
Overall	**275**	**161**	**16**	**98**	**41.2**	**35.6**

4.5.2 Enhancing understanding of rules through alternative representations

As part of the experiment, we also sought to investigate the effect of providing alternative presentations of the rule to see if that affected participants' understanding of the rule measured in the paper-based task. This was done for the CSP-3 algorithm treatments (C3, C3-I and C3-I2), given the additional complexity of that algorithm and the department's use of it as part of the Compliance-Based Inspection System.

Half the participants in the CSP-3 rule treatments were randomly assigned to receive a diagrammatic representation of the inspection rule they experienced in the main experimental task.[56] Figures 20 and 21 in Appendix B show the diagrams received by

[56] To assess whether prior exposure to the CSP-3 algorithm in the main experimental task also improved understanding of the rule, we gave participants in the boundary treatments (M and R)

those subjects, noting that the diagrams refer to "consignments" rather than references to "goods" as used elsewhere in the experiment.[57] The CSP-3 rule diagram, which was a simplified version of a diagram that the department had been using for internal communications, was given alongside the paper-based tasks after they had completed the computer-based tasks, including the post-experiment questionnaire.

Table 17 below shows the performance on the paper-based task and shows almost no difference between the two groups for the CSP-3 treatments.[58] However, such a comparison does not control for other factors that may affect performance on the task, such as prior experience working with technical rules or exposure to concepts from probability through university study.[59]

Table 17: Performance on paper-based task for CSP-3 treatments

Received CSP-3 diagram	Scenarios correct in paper-based task			
	None	One	Two	Total
Yes	14 (48.3)	5 (17.2)	10 (34.5)	29
No	13 (46.4)	5 (17.9)	10 (35.7)	28
Total	27	10	20	57

Notes: Results based on subjects experiencing treatments C3, C3-I and C3-I2 only. Figures in parentheses below the counts are percentages based on row totals.

the CSP-3 algorithm with full information in their paper-based task. Those in treatment M received the diagram, while subjects in treatment R did not. Significant differences from either a statistical or an economic perspective were not evident between these treatments and those who experienced the CSP-3 algorithm as part of their main experimental task. This is not surprising, as relatively few subjects – only 11 in total – received either treatments M or R.

[57] We thank one of the anonymous scientific reviewers for pointing out this discrepancy in wording and acknowledge that this discrepancy may have affected the usefulness of the diagram in helping subjects understand the CSP-3 algorithm.

[58] The p-value assumed with the chi-squared test of independence exceeds 0.8. This finding should also be considered in light of the wording discrepancy highlighted in the previous footnote.

[59] Casual observation during the experimental sessions suggested that the understanding of rules might be related to a subject's field of study. In particular, it appeared anecdotally that engineering students, who constituted around one-quarter of the experimental subjects, may have performed better in paper-based task than students from other fields. The statistical evidence was limited in this regard. In the CSP-3 treatments, engineering students performed slightly worse in the paper-based task than students from other majors; only 23.1 per cent of engineering students solved both test questions compared to 38.6 per cent for other students. Across all treatments, engineering students performed slightly better (41.7 per cent answered both task questions correctly) than their non-engineering counterparts (33.5 per cent).

5. Supplier Choices and the Influence of Treatment Effects

In this chapter, we investigate the impact of the 18 treatments used in the experiment, which were summarised in Table 2 of the *Final Report*. These treatments vary across a number of different dimensions. To better understand the impact of each dimension on supplier choice, we compare supplier choices in the relevant treatments, and investigate how particular features of interest affect importer decisions that have implications for the design of regulation seeking to manage biosecurity risks at the border.

Before doing so, we outline the approach used to analyse the experimental data in a way that seeks to make valid and robust inferences about the causal impacts of the specific changes in the key aspects assessed in the experiment.

5.1 Empirical strategy to make valid inferences about treatment effects

While it is tempting to compare differences between treatments through the implied approach rates of biosecurity risk material, such comparisons are likely to yield biased estimates of the "true" influence of the different rules on supplier choices. As we illustrate in Chapter 6, differences in individual characteristics of participants can lead subjects to behave in markedly different ways in response to the same sets of rules. Given the small number of participants in each treatment, it would be courageous to assume the distribution of these characteristics, some of which cannot be easily observed, is identical across all 18 treatment groups.

Given we observe individuals making repeated choices over time,[60] observed patterns of decision-making for subjects may change over time as they learn the rules and procedures. Hence, in assessing whether differences in observed behaviour that can be attributed to changes in inspection rules, we seek to account for period effects (in the time-series dimension of our data) and differences in the individual characteristics of participants in each treatment (in the cross-sectional dimension of our data).

To estimate the role of each of the factors we varied in the experimental design, we estimate a series of linear regression models, where the implied approach rate of biosecurity risk material associated with each supplier choice is the dependent variable.[61] For explanatory variables in these statistical models, we use dummy variables to measure treatment differences, together with measures of individual characteristics elicited from the other experimental tasks, including the post-experiment questionnaire, and control for period effects. These regression models are used to determine whether differences between pairwise-comparable treatments are significant from a statistical and/or practical viewpoint. They are

[60] This feature, which implies both cross-sectional and time-series dimensions the structure of our experimental data, means our statistical frameworks for analysing the data fall into the category referred to as *panel-data models*. For a detailed discussion of the statistical models appropriate for analysing this type of data, see Hsiao (2014).

[61] As a robustness check, we also estimated ordered probit models with the supplier chosen as dependent variable. The implied behavioural patterns are very similar to the random-effects model estimates presented here with the approach rate as the dependent variable.

estimated as random effects panel-data models[62] using feasible generalised least squares.[63]

In summarising our statistical results, we always present two models for each comparison of treatments. The first model only controls for periods effects, whereas the second model in the pair incorporates additional individual-level control variables covering:

- demographic factors, specifically gender and nationality (via a dummy variable taking the value one if the subject is not Australian);
- the choice of gamble in the lottery-choice task (Task 1); and
- the participant characteristics from the post-experimental questionnaire around their attitudes towards environmental concerns, government and politics, and their understanding of the inspection rule.

Before proceeding to describe the results from these statistical models, we note a few aspects that suggest caution in interpreting the results. Overall, we find relatively few treatment effects are statistically different from zero at the "conventional" levels of significance.[64] The main reasons why the treatment effects are not significant, in the main, are the relatively small sample sizes in each treatment group and the substantial heterogeneity in responses within the treatment groups. These features of the experiment also imply that statistical hypothesis tests likely suffer from low power, in that there is limited ability to distinguish "true" non-zero treatment effects. Consequently, standard errors of the estimated treatment effects therefore tend to be relatively large, implying that what appear to be sizeable treatment effects cannot be distinguished from there being no treatment effect in a statistical sense. Repeating the experiment with two to four times the number of subjects per treatment could easily lower the standard errors of the estimated treatment effects to an extent that many would be deemed "statistically significant". Furthermore, the ability to correctly distinguish "true" non-zero treatment effects would also improve.

As highlighted in the *Final Report*, while the large number of treatments enabled the project team to investigate a range of issues around the inspection process that were of practical relevance to the department, such an approach has drawbacks from a statistical perspective. By assessing so many different pairwise treatment

---

62 The random-effects specification is appropriate in this case as the individuals included in the subject pool are drawn from a larger population and we are willing to assume that the unobserved individual effects are uncorrelated with the included variables. Details of the random-effects model specification and estimation procedure can be found in many standard econometrics textbooks; see, for example, Chapter 13.4 of Greene (2003).

63 As a further robustness check, we also estimated these panel-data models using as pooled regression models, where the data are pooled across individuals and time periods. In doing so, we correct the standard errors for the coefficient estimates for potential heteroscedasticity between individual participants and as well as clustering effects, given that supplier selections made by the same subject are unlikely to be statistically independent. Those results are very similar to the preferred model specification presented in this report, but are available upon request from the authors.

64 The "conventional" thresholds of statistical significance for hypothesis testing (at the 10, 5 and 1 per cent levels) may not be "appropriate" in all cases. Instead, the choice of an appropriate level of significance for testing hypotheses in the frequentist statistical paradigm depends on the relative costs of potential decision-errors (Type I and Type II errors). Drawing on this decision-theoretic framework may mean the appropriate level of significance may well be higher (or lower) than those employed for reporting purposes in this report.

comparisons, the analysis is subject to multiple testing problems, as highlighted by List et al. (2016). We have chosen not to adopt the types of corrections suggested in that paper for simplicity of exposition; however, in doing so, we are cognisant that this enhances the likelihood of identifying treatment effects exist when in fact they may not be present.

Finally, some care needs to be taken in interpreting the findings in models where individual-level controls have been added. These models run the risk of being over-specified, as the number of individuals per treatment group is comparatively low. In some cases, we needed to drop some of the nominated control variables from the regression models to avoid issues associated with multicollinearity, such as obviously inflated standard errors.

The remainder of this chapter focuses on differences in behaviour that can be attributed to differences in the rules changed by the experimenter; the effect of the individual characteristics on supplier choices is the focus of Chapter 6.

5.2 Different inspection rules

We investigate two types of compliance-based inspection rules, which use the CSP-1 and CSP-3 algorithms, in this experiment. These two rules differ slightly differ in their response to failing an inspection when in the monitoring phase of the algorithm, with the CSP-3 algorithm being slightly more forgiving than the CSP-1 algorithm in the event of an inspection failure. In this section, we compare these rules under circumstances where the participants have complete information about the rule (treatments C1 and C3) and where there is incomplete information about the probability of inspection in monitoring mode (treatments C1-I and C3-I). This section is to be read in conjunction with Chapter 4.2 of the *Final Report*.

5.2.1 Regression modelling of the implied approach rate

Model 1 in Table 18 shows that the difference in the implied approach rate is not statistically different from zero once period effects are accounted for.[65] With additional individual-level control variables, model 2 suggests that the implied approach rate of biosecurity risk material is 4.2 percentage points lower in treatment C3 and in treatment C1 – a result that is statistically different from zero at the 5 per cent level of significance.[66] Model 2 also shows that two control variables, the nationality dummy and the variable capturing participants' understanding of the inspection rule, are important factors influencing supplier choice in this comparison of treatments.

Turning to an environment where importers do not know the probability of inspection in monitoring mode, Table 7 in the *Final Report* shows that the average implied approach rate for biosecurity risk material is almost identical in treatments C1-I and C3-I. Models 3 and 4 in Table 18 suggest that there are no significant differences between the incomplete-information treatments; if anything, there is some tentative

[65] The p-value for this test of hypothesis for model 1 in Table 18 is 0.203, with an associated 95 per cent confidence interval for the treatment effect of [-0.061,0.013].

[66] The p-value for this hypothesis test is 0.024 and the associated 95 per cent confidence interval for the treatment effect in model 2 in Table 18 is [0.006,0.078].

evidence to suggest the approach rate could be slightly higher in treatment C3-I than in treatment C1-I after controlling for other factors.[67]

Table 18: Treatment difference models reflecting different inspection rules

Model	(1)	(2)	(3)	(4)
Dependent variable	Implied biosecurity risk material approach rate			
CSP-3 rule	-0.024 (0.019)	-0.042** (0.019)	0.004 (0.032)	0.040 (0.032)
Non-Australian		0.059** (0.026)		0.106** (0.043)
Male		-0.029 (0.020)		-0.071* (0.041)
Gamble choice in Task 1		0.003 (0.006)		0.019* (0.010)
Belief in successful government intervention		0.008 (0.010)		0.027* (0.015)
Relative importance of environment/economy		∂		-0.010 (0.012)
Concerns for pests and diseases		-0.004 (0.013)		0.035 (0.022)
Extinction concerns		0.012 (0.010)		0.009 (0.026)
Level of conservatism		-0.001 (0.005)		-0.000 (0.008)
Rule understanding		-0.007* (0.004)		-0.000 (0.005)
Number of observed choices	2200	2200	1500	1500
Subjects	44	44	30	30
Comparison treatment	C1	C1	C1-I	C1-I

Notes: *, ** and *** indicate that the relevant coefficient is statistically different from zero at the 10, 5 and 1 per cent level of significance respectively. Figures in parentheses refer to standard errors of the estimated model coefficients. Estimates of the coefficients for the period controls and constant term have been suppressed. Models estimated using feasible generalised least squares with the "standard" random-effects estimator. ∂ indicates a control variable that was dropped due to the presence of multicollinearity when all controls were included; refer to the discussion in Chapter 5.1 for the reason behind this procedure.

[67] The respective p-values and confidence intervals for the treatment effects in Table 18 are:
- model 3 – p-value of 0.896 and 95 per cent confidence interval of [-0.058,0.066]; and
- model 4 – p-value of 0.220 and 95 per cent confidence interval of [-0.024,0.103].

5.2.2 Loss-function approach to comparing different inspection rules

In Chapter 2.2.6, we discussed the possibility of using metrics other than the implied approach rate to assess the performance of different rules from the regulator's perspective. We now illustrate the idea of using a loss function to illustrate how comparisons based on different criteria can affect conclusions around which rule would be preferred by the regulator. To parameterise the loss function, the cost of undertaking inspections is set at one unit to function as a numeraire. The implied cost of leakage from contaminated goods not inspected is then varied to map various scenarios.

For exposition purposes, we compare the experimental outcomes of treatments C1 and C3 (respectively, the CSP-1 and CSP-3 treatments with full information). The comparison is based on the raw experimental data, without conditioning on other factors that may influence participant behaviour, such as those included in model 2 in Table 18.

From a theoretical perspective, an experimental subject following the CSP-3 algorithm should receive fewer inspections than one following the CSP-1 algorithm, all other things being equal. This is confirmed in the experimental data, as shown in the left-hand panel of Figure 15, though the tendency for participants in treatment C3 to choose suppliers with lower biosecurity risk material approach rates also contributes to receiving fewer inspections. For treatment C3, the median number of inspections performed during the experiment was 238 (out of a total of 500 goods imported during the experiment), while the median number of inspections for treatment C1 was 280. The left-hand panel of Figure 15 also highlights the heterogeneity of outcomes in both treatment groups, particularly in treatment C1.

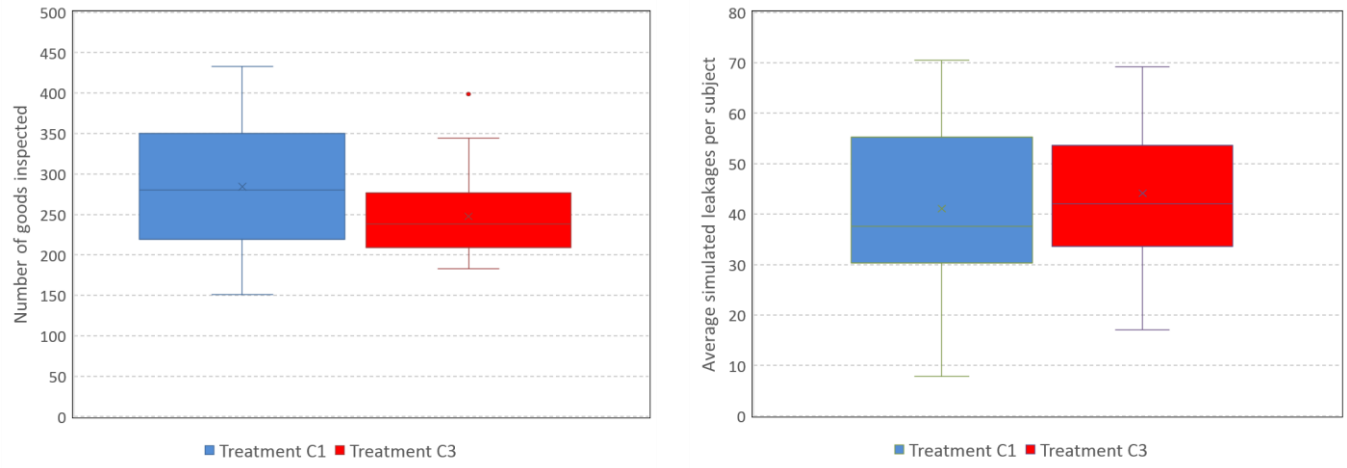


Figure 15: Distribution of inspections (left-hand panel) and implied distribution of leakages (right-hand panel) in treatments C1 and C3

In implementing the loss-function approach, a measure of leakages also must be developed. For this experimental setup, leakages of biosecurity risk material are not observed directly as part of the experimental output; instead, they must be simulated from goods not selected for inspection using the approach rate for the supplier chosen

by the importer.[68] The right-hand panel of Figure 15 indicates the distribution of the mean number of leakages per experimental subject implied by their supplier choices for non-inspected goods, where the average is computed over 500 simulations. Treatment C1 appears to have slightly fewer leakages "on average" than treatment C3,[69] though the results confirm the significant heterogeneity of behaviours demonstrated in the experiment. The notion that the CSP-1 algorithm would admit fewer leakages than the CSP-3 algorithm, with all other things held constant, reflects the CSP-1 algorithm's less forgiving penalty structure for failing inspections and aligns with theoretical predictions apparent from the rules' respective Markov chains.[70]

As the experimental results and follow-up simulations suggest a trade-off between the number of inspections performed and the implied rate of leakage of biosecurity risk material, the regulator's preferred rule will depend on the relative costs they perceive for leakage versus undertaking an inspection. Figure 16 plots the regulator's loss function associated with treatments C1 and C3 across a range of values for the relative cost of leakage. It suggests that if the relative cost of leakage is small, then the CSP-3 algorithm would be preferable as the lower costs associated with fewer inspections tend to dominate the slightly higher rate of leakage admitted. However, if the regulator believes that the cost to them of leakage exceeds around 12 times the cost of undertaking inspection, then the preferred rule switches to the CSP-1 algorithm. As the relative cost of leakage rises, the loss function associated with the CSP-1 algorithm increases more slowly than that for the CSP-3 algorithm, widening the gap between these two alternative rules. This reinforces recommendations made in Rossiter and Hester (2017) that the CSP-1 algorithm is preferred by the regulator when the consequences of biosecurity risk material leakage are "high".

[68] As the inspection process has been assumed to be "perfect", as discussed in Chapter 3.1, leakages cannot occur from goods that were chosen for inspection by the biosecurity regulator. Another implicit assumption here is that the regulator also "knows" the approach rate of the different suppliers to compute the loss function.

[69] The simulation results indicate, under either rule, that the rate of post-intervention compliance is around 91 to 92 per cent. It is worth noting that a pathway exhibiting these types of relatively low rates of compliance would be highly unlikely to be considered eligible to be added to the department's Compliance-Based Inspection Scheme.

[70] See Part C of the Online Appendix of Rossiter and Hester (2017) for an illustration of these properties.

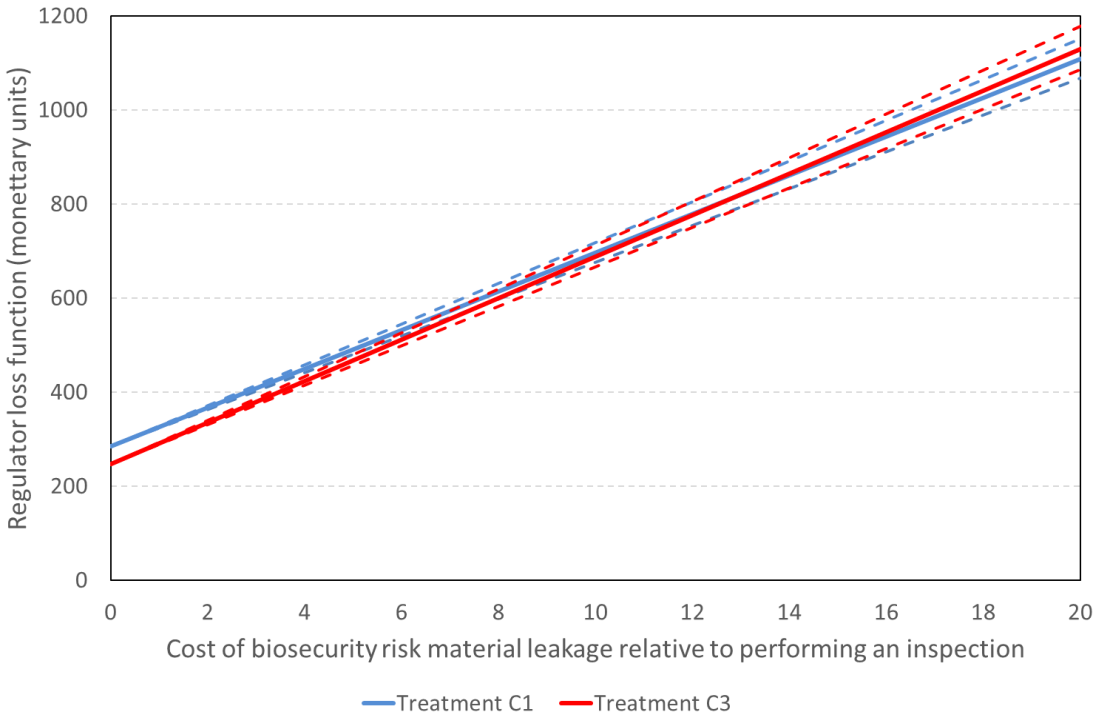


Figure 16: Regulator loss function under treatments C1 and C3 for different relative costs of leakage versus undertaking inspections

Notes: Dashed lines represent 95 per cent distributional bounds on the implied post-intervention leakage of biosecurity risk material. These are designed to reflect the simulated nature of information on leakage.

5.3 Level of information about the rule

We now turn to investigate the role of providing different levels of information on the rule structure and parameters on the choices of supplier made by experimental subjects. The comparisons presented here focus on the three treatments involving the CSP-3 algorithm. Treatment C3 provides the experimental subjects with the full rule specification, while treatment C3-I provides "vague" information on the monitoring fraction in terms of a range. Treatment C3-I2, which most closely resembles the department's practice at the time the experiments were conducted, provides a range for the monitoring fraction and only vaguely describes the tight census number used in the CSP-3 algorithm. This section is to be read in conjunction with Chapter 4.3 of the *Final Report*.

A priori, it seems plausible to assume that the approach rate for biosecurity risk material in the earlier periods of the main experimental task could be higher in treatment C3 than in C3-I. This is because the "true" monitoring fraction for both treatments is 0.2, while treatment C3-I describes the monitoring fraction as between 0.1 and 0.5. During the course of the experiment, participants may figure out that the actual inspection probability in C3-I is around 0.2, potentially resulting in differences in observed supplier choices between these two treatments narrowing as the task is completed. Interestingly, this trend is not what we observe. The *Final Report* highlights that the implied approach rate is lowest in the full information treatment (treatment C3) and highest in treatment C3-I where the monitoring fraction is vaguely described.

Table 19 provides evidence for the statistical strength of these differences. As shown in models 1 and 2, the implied approach rate is higher in treatment C3-I than in treatment C3 by 3.1 to 5.0 percentage points. While the treatment effect is not statistically different from zero if only period effects are accounted for (model 1), model 2 indicates that providing less information to experimental subjects about the rule parameters is associated with them choosing higher-risk suppliers on average.[71]

Table 19: Treatment difference models reflecting the level of information about the rule

Model	(1)	(2)	(3)	(4)
Dependent variable	Implied biosecurity risk material approach rate			
C3-I	0.031 (0.022)	0.050** (0.024)		
C3-I2			0.021 (0.017)	0.023 (0.018)
Non-Australian		0.049 (0.031)		0.024 (0.030)
Male		-0.014 (0.026)		-0.030* (0.018)
Gamble choice in Task 1		0.009 (0.007)		-0.004 (0.006)
Belief in successful government intervention		0.026** (0.012)		0.004 (0.012)
Relative importance of environment/economy		-0.003 (0.011)		0.003 (0.009)
Concerns for pests and diseases		0.017 (0.016)		-0.011 (0.015)
Extinction concerns		-0.005 (0.020)		0.005 (0.015)
Level of conservatism		-0.002 (0.005)		0.002 (0.005)
Rule understanding		-0.006* (0.004)		-0.002 (0.003)
Number of observed choices	2050	2050	1950	1950
Subjects	41	41	39	39
Comparison treatment	C3	C3	C3	C3

Notes: *, ** and *** indicate that the relevant coefficient is statistically different from zero at the 10, 5 and 1 per cent level of significance respectively. Figures in parentheses refer to standard errors of the estimated model coefficients. Estimates of the coefficients for the period controls and constant term have been suppressed. Models estimated using feasible generalised least squares with the "standard" random-effects estimator.

[71] The respective p-values and 95 per cent confidence intervals for the treatment effects in Table 19 are:

- model 1 – p-value of 0.157 and 95 per cent confidence interval of [-0.074,0.012]; and
- model 2 – p-value of 0.036 and 95 per cent confidence interval of [0.003,0.096].

Table 19 also considers the performance of treatment C3-I2, which mirrors the form of rule currently used by the department, relative to treatment C3. *A priori*, it is unclear whether the additional layer of incomplete information lowers or increases the implied biosecurity risk material approach rate as compared to treatment C3-I. Our results suggest that the biosecurity risk material approach rate implied by the supplier choices for treatment C3-I2 lies between treatments C3 and C3-I for most periods. In particular, the *Final Report* shows that the implied approach rate is higher in treatment C3-I2 than in treatment C3 in all periods except the final period of the task. Models 3 and 4 in Table 19 show that the implied approach rate is higher in treatment C3-I2 than the full-information treatment (treatment C3), but the difference is not significant from a statistical perspective.[72]

5.4 Framing feedback on rule performance

We study the role of framing in the feedback provided to our experimental subjects by including a gain frame (treatments C1-G and C1-IG) and a loss frame (treatments C1-L and C1-IL) in our treatments to compare with the CSP-1 treatments with the fully specified rule (treatment C1) and the rule with the monitoring fraction vaguely described (treatment C1-I). The gain frame specified the amount saved from not being inspected in the feedback given to the participants after each period. The loss frame specified the costs (that is, monetary losses) incurred by the experimental subject due to their consignments being inspected. In both cases, experimental subjects received this additional feedback from an additional statement on the results screen that highlighted these particular performance measures. This section is to be read in conjunction with Chapter 4.4 of the *Final Report.*

Our conjecture in this experiment was that both frames of feedback could result in lower implied approach rates of biosecurity risk material. As we discuss in this section, the results accord with our prior intuition, with some economically sizeable effects present; however, the impact of framing tends to be imprecisely measured in the statistical models presented here and thus the impact of framing is not statistically significant.

5.4.1 Impact of the gain frame

The *Final Report* shows that the implied approach rate is lower in the gain-frame treatment under complete rule specification (treatment C1-G) than for the baseline treatment C1. However, the regression analysis in Table 20 shows that the treatment differences are small in magnitude and not statistically different from zero. This holds when only period effects are applied (model 1) and when individual-level controls are also included (model 2).[73]

---

[72] The p-values for statistical tests and 95 per cent confidence intervals associated with the treatment effect for models 3 and 4 in Table 19 are:
- model 3 – p-value of 0.209 and 95 per cent confidence interval of [-0.012,0.054]; and
- model 4 – p-value of 0.202 and 95 per cent confidence interval of [-0.013,0.059].

[73] The respective p-values and 95 per cent confidence intervals for the treatment effects in Table 20 are:
- model 1 – p-value of 0.660 and 95 per cent confidence interval of [-0.066,0.042]; and
- model 2 – p-value of 0.894 and 95 per cent confidence interval of [-0.049,0.056].

Table 20: Treatment difference models reflecting differences in feedback from the gain frame

Model	(1)	(2)	(3)	(4)
Dependent variable	Implied biosecurity risk material approach rate			
Gain frame	-0.012 (0.028)	0.004 (0.027)	-0.035 (0.026)	-0.032 (0.030)
Non-Australian		0.094*** (0.030)		0.023 (0.048)
Male		-0.041 (0.029)		-0.003 (0.034)
Gamble choice in Task 1		0.006 (0.009)		-0.000 (0.012)
Belief in successful government intervention		∂		∂
Relative importance of environment/economy		-0.018 (0.012)		-0.008 (0.012)
Concerns for pests and diseases		-0.005 (0.018)		0.018 (0.019)
Extinction concerns		0.008 (0.014)		0.006 (0.019)
Level of conservatism		0.003 (0.007)		0.004 (0.007)
Rule understanding		-0.007 (0.005)		0.003 (0.005)
Number of observed choices	1950	1950	1500	1500
Subjects	39	39	30	30
Comparison treatment	C1	C1	C1-I	C1-I

Notes: *, ** and *** indicate that the relevant coefficient is statistically different from zero at the 10, 5 and 1 per cent level of significance respectively. Figures in parentheses refer to standard errors of the estimated model coefficients. Estimates of the coefficients for the period controls and constant term have been suppressed. Models estimated using feasible generalised least squares with the "standard" random-effects estimator. ∂ indicates a control variable that was dropped due to the presence of multicollinearity when all controls were included; refer to the discussion in Chapter 5.1 for the reason behind this procedure.

Turning to the treatments where the monitoring fraction is specified as a range, we find that the impact of the gain frame becomes more pronounced. Model 3 highlights that the implied average approach rate is 3.5 percentage points lower for treatment C1-IG than for treatment C1-I. Furthermore, the *Final Report* shows that the implied approach rate is always lower across all periods if a gain frame is included, with the right-hand panel of Figure 5 (*Final Report*, page 46) suggesting that the confidence bands around the implied approach rate estimates do not overlap for most of the "middle periods" of the experimental task.

When comparing the results across all 50 periods of the experiments, regression models 3 and 4 in Table 20 suggest that the estimated reduction in the approach rate

could be economically significant, at between 3.2 and 3.5 percentage points, but the treatment difference is not statistically different from zero at the usual levels of significance.[74] This probably reflects the high degree of heterogeneity between individual choice patterns under these treatments.

As an illustrative check of the robustness of findings, we re-estimate models 3 and 4 using information based on the middle 30 periods of the task (that is, excluding the first ten and last ten periods of the task). The results for this assessment are shown in Table 21, with the models based on the sub-samples referred to as models 3A and 4A. Somewhat at odds with the graphical presentation of the results in the *Final Report*, the statistical insignificance of the treatment effect either without (model 3A) or with individual-level controls (model 4A) appears to be confirmed when confining consideration to this subset of subject choices.[75]

[74] The respective p-values and 95 per cent confidence intervals for the treatment effects in Table 20 are:

- model 3 – p-value of 0.174 and 95 per cent confidence interval of [-0.085,0.015]; and
- model 4 – p-value of 0.281 and 95 per cent confidence interval of [-0.090,0.026].

[75] The respective p-values and 95 per cent confidence intervals for the treatment effects in Table 21 are:

- model 3A – p-value of 0.178 and 95 per cent confidence interval of [-0.094,0.017]; and
- model 4A – p-value of 0.305 and 95 per cent confidence interval of [-0.097,0.030].

Table 21: Robustness check for treatment difference models for the gain frame under incomplete information about the rule

Model	(3)	(3A)	(4)	(4A)
Dependent variable	Implied biosecurity risk material approach rate			
Gain frame	-0.035 (0.026)	-0.038 (0.029)	-0.032 (0.030)	-0.033 (0.032)
Non-Australian			0.023 (0.048)	0.026 (0.052)
Male			-0.003 (0.034)	0.003 (0.037)
Gamble choice in Task 1			-0.000 (0.012)	0.004 (0.014)
Belief in successful government intervention			∂	∂
Relative importance of environment/economy			-0.008 (0.012)	-0.011 (0.013)
Concerns for pests and diseases			0.018 (0.019)	0.027 (0.021)
Extinction concerns			0.006 (0.019)	0.006 (0.021)
Level of conservatism			0.004 (0.007)	0.004 (0.008)
Rule understanding			0.003 (0.005)	0.003 (0.006)
Number of observed choices	1500	900	1500	900
Subjects	30	30	30	30
Comparison treatment	C1-I	C1-I	C1-I	C1-I

Notes: *, ** and *** indicate that the relevant coefficient is statistically different from zero at the 10, 5 and 1 per cent level of significance respectively. Figures in parentheses refer to standard errors of the estimated model coefficients. Estimates of the coefficients for the period controls and constant term have been suppressed. Models estimated using feasible generalised least squares with the "standard" random-effects estimator. ∂ indicates a control variable that was dropped due to the presence of multicollinearity when all controls were included; refer to the discussion in Chapter 5.1 for the reason behind this procedure.

5.4.2 Impact of the loss frame

For the situation where subjects know the full rule specification, models 1 and 2 in Table 22 confirm that the implied approach rate is lower on average (by between 2.2 and 3.8 percentage points) under the loss-frame treatment (treatment C1-L) than

both the baseline CSP-1 rule treatment (treatment C1).[76] However, the estimated coefficients on the gain frame dummy variable are not statistically different from zero under conventional hypothesis-testing procedures.[77]

Table 22: Treatment difference models reflecting differences in feedback from the loss frame

Model	(1)	(2)	(3)	(4)
Dependent variable	Implied biosecurity risk material approach rate			
Loss frame	-0.022 (0.024)	-0.038 (0.025)	-0.002 (0.024)	-0.026 (0.035)
Non-Australian		0.069** (0.031)		-0.039 (0.052)
Male		-0.028 (0.027)		-0.032 (0.036)
Gamble choice in Task 1		0.011 (0.008)		0.011 (0.010)
Belief in successful government intervention		0.004 (0.017)		0.009 (0.017)
Relative importance of environment/economy				-0.007 (0.013)
Concerns for pests and diseases		-0.008 (0.015)		0.028* (0.016)
Extinction concerns		0.026** (0.012)		0.004 (0.018)
Level of conservatism		0.004 (0.006)		-0.003 (0.008)
Rule understanding		-0.000 (0.004)		0.003 (0.005)
Number of observed choices	1900	1900	1450	1450
Subjects	38	38	29	29
Comparison treatment	C1	C1	C1-I	C1-I

Notes: *, ** and *** indicate that the relevant coefficient is statistically different from zero at the 10, 5 and 1 per cent level of significance respectively. Figures in parentheses refer to standard errors of the estimated model coefficients. Estimates of the coefficients for the period controls and constant term have been suppressed. Models estimated using feasible generalised least squares with the "standard" random-effects estimator. ∂ indicates a control variable that was dropped due to the presence of multicollinearity when all controls were included refer to the discussion in Chapter 5.1 for the reason behind this procedure.

[76] The *Final Report* also shows that the implied approach rate for the loss-frame treatment under full rule specification (treatment C1-L) is lower than the corresponding gain-frame treatment (treatment C1-G).

[77] The respective p-values and 95 per cent confidence intervals for the treatment effects in Table 22 are:

- model 1 – p-value of 0.349 and 95 per cent confidence interval of [-0.069,0.024]; and
- model 2 – p-value of 0.131 and 95 per cent confidence interval of [-0.087,0.011].

Interestingly, the loss frame is less effective in the environment where there is uncertainty for the subject about their monitoring fraction. Models 3 and 4 in Table 22 show weak evidence of a reduction in biosecurity risk material approach rates under the loss frame; these effects are measured imprecisely and the coefficients attached to the loss frame dummy variable are not significantly different from zero.[78]

5.5 Costs of being inspected and failing inspection

These comparisons for the key cost parameters associated with inspection and treatment seek to confirm the theoretical predictions, consistent with the findings of Rossiter and Hester (2017) and the calibration exercise in Chapter 3, that higher costs of being inspected and/or failing inspection encourage the choice of lower-risk suppliers. Chapter 4.5 in the *Final Report* suggests that the direction of treatments effects align with the theoretical predictions and that higher inspection costs have a limited impact on improving compliance when treatment costs are already high.

Table 23 provides two models showing the impact of cost changes on the implied approach rate associated with supplier choices. The comparison treatment in these regressions is treatment C1-4.12.

We observe that reducing treatment costs to 6 monetary units (treatment C1) tends to increase the approach rate of biosecurity risk material by 3.1 to 3.3 percentage points, though this impact is not statistically significant.[79] While reducing inspection costs to 2 monetary units has no significant impact on the biosecurity risk material approach rate when treatment costs remain at 12 monetary units,[80] reducing both inspection and treatment costs (treatment C1-2.6) leads to a very large and statistically significant increase in the approach rate of biosecurity risk material.[81]

Overall, the results from this section confirm our intuition and the theoretical predictions that compliance-based inspection protocols are likely to be most appropriate for plant-based products where the cost of failing inspection is high or, to a lesser extent, where the costs associated with being inspected are high.

---

78 The respective p-values and 95 per cent confidence intervals for the treatment effects in Table 22 are:
- model 3 – p-value of 0.935 and 95 per cent confidence interval of [-0.048,0.045]; and
- model 4 – p-value of 0.454 and 95 per cent confidence interval of [-0.094,0.042].

79 The respective p-values and 95 per cent confidence intervals associated with the coefficient attached to treatment C1 (compared to treatment C1-4.12) in models 1 and 2 of Table 23 are:
- model 1 – p-value of 0.209 and 95 per cent confidence interval of [-0.019,0.085]; and
- model 2 – p-value of 0.480 and 95 per cent confidence interval of [-0.033,0.069].

80 The respective p-values and 95 per cent confidence intervals associated with the coefficient attached to treatment C1-2.12 (compared to treatment C1-4.12) in models 1 and 2 of Table 23 are:
- model 1 – p-value of 0.900 and 95 per cent confidence interval of [-0.054,0.061]; and
- model 2 – p-value of 0.589 and 95 per cent confidence interval of [-0.072,0.041].

Note too that this coefficient changes sign with the inclusion of individual-level control variables.

81 The respective p-values and 95 per cent confidence intervals associated with the coefficient attached to treatment C1-2.6 (compared to treatment C1-4.12) in models 1 and 2 of Table 23 are:
- model 1 – p-value of 0.004 and 95 per cent confidence interval of [0.028,0.142]; and
- model 2 – p-value of 0.008 and 95 per cent confidence interval of [0.019,0.127].

Table 23: Treatment difference models reflecting different inspection and treatment cost parameters

Model	(1)	(2)
Dependent variable	Implied biosecurity risk material approach rate	
C1	0.033 (0.026)	0.018 (0.026)
C1-2.6	0.085*** (0.029)	0.073*** (0.028)
C1-2.12	0.004 (0.029)	-0.016 (0.029)
Non-Australian		0.045** (0.020)
Male		0.008 (0.018)
Gamble choice in Task 1		0.006 (0.005)
Belief in successful government intervention		∂
Relative importance of environment/economy		∂
Concerns for pests and diseases		∂
Extinction concerns		∂
Level of conservatism		∂
Rule understanding		-0.004 (0.003)
Number of observed choices	2700	2700
Subjects	54	54
Comparison treatment	C1-4.12	C1-4.12

Notes: *, ** and *** indicate that the relevant coefficient is statistically different from zero at the 10, 5 and 1 per cent level of significance respectively. Figures in parentheses refer to standard errors of the estimated model coefficients. Estimates of the coefficients for the period controls and constant term have been suppressed. Models estimated using feasible generalised least squares with the "standard" random-effects estimator. ∂ indicates a control variable that was dropped due to the presence of multicollinearity when all controls were included; refer to the discussion in Chapter 5.1 for the reason behind this procedure.

5.6 Regulatory environment with a choice of inspection rule

As part of designing a simple menu of regulatory contracts to understand the role that rule choice could play in influencing supplier choices, we constructed a second CSP-1 rule with different parameters. As discussed in Chapter 3.2 of the *Final Report*, these mechanisms took the parameter values:

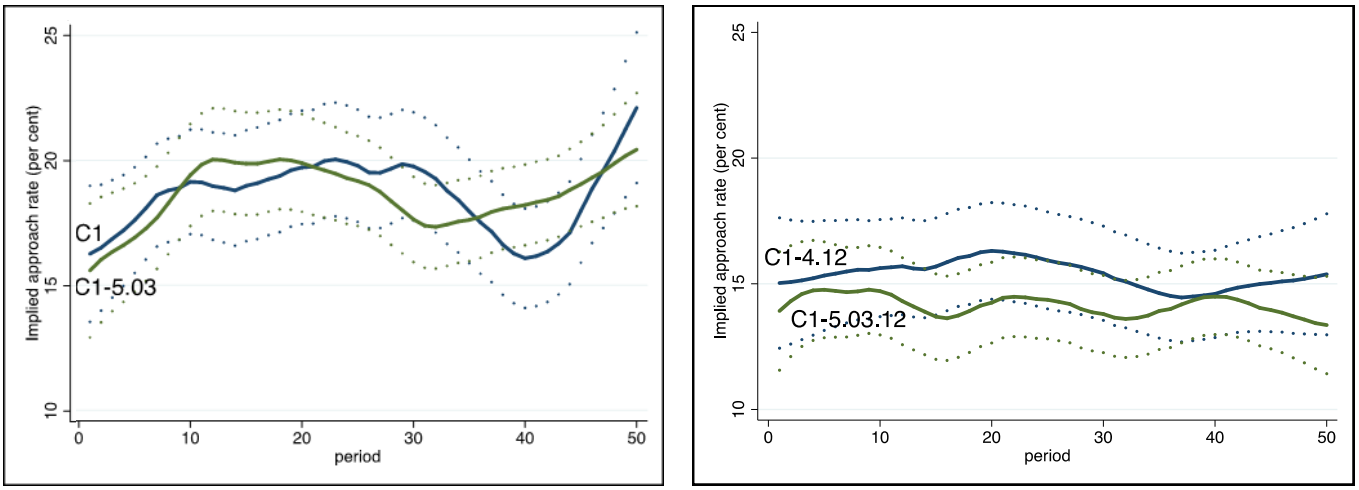
- Mechanism A (the standard rule): clearance number 10 and monitoring fraction 0.2; and
- Mechanism B: clearance number 5 and monitoring fraction 0.3.

We compare the two mechanisms, both with standard (6) and high (12) treatment costs. Before we investigate the role of choice, we compare behaviour in these two

regimes when individuals are "forced" to make decisions in these regimes. However, note that the direct comparison between these two regimes is difficult to interpret as the regimes differ on two dimensions simultaneously. This is particularly tricky in this context, given the two effects should partly offset each other.

Under standard treatment costs, the theoretical predictions for a risk-neutral experimental subject (see Table 7 in the *Final Report*) suggest that the implied average approach rate should be higher under Mechanism B (treatment C1-5.03) than Mechanism A (treatment C1). However, the results in the left-hand panel of Figure 17 suggest that the biosecurity risk material approach rate implied by subjects' supplier choices is almost identical in these two treatments.[82]

If we compare these two mechanisms with high treatment costs (that is, treatments C1-4.12 and C1-5.03.12), we again observe relatively small differences in the implied approach rate stemming from subjects' supplier choices (see the right-hand panel of Figure 17). This accords with the theoretical prediction for these two treatment comparisons, where the "optimal" behaviour of a risk-neutral subject should not differ between these treatments. We find the average approach rate of biosecurity risk material is somewhat higher in treatment C1-4.12 than in treatment C1-5.03.12 – a pattern that seems to hold across all periods.[83] However, the 95 per cent confidence interval bands are very wide, reflecting high individual variance in the treatments and the small number of participants (nine) who received treatment C1-4.12.



Standard treatment cost comparisons High treatment cost comparisons

Figure 17: Comparison of implied biosecurity risk material approach rates of CSP-1 rules (Mechanisms A and B) under standard and high treatment costs

Now we investigate the role of choice. In treatments Choice6 and Choice12, we gave participants the possibility to choose whether they wished to follow the inspection

82 Statistical models of the kind considered elsewhere in this chapter confirmed that there was virtually no difference between the treatments once period effects or period and individual-level effects were controlled for.

83 Regression analysis confirms that the difference in average implied approach rates is not statistically significant after controlling for period effects. However, panel regressions including controls for individual-level and period effects suggests that the approach rate of biosecurity risk material is marginally lower in treatment C1-5.03.12. We place little weight on this finding, which may be an artefact of the low number of independent observations in each treatment group.

rule given by Mechanism A or Mechanism B. More precisely, in treatment Choice6 they could choose between treatments C1 and C1-5.03. Similarly, in treatment Choice12 they could choose between treatments C1-4.12 and C1-5.03.12, which correspond to the previous two treatments but incur a treatment cost of 12 monetary units rather than 6 monetary units.

Previous research in the experimental economics literature on choice supports the behavioural economics explanation that offering choice could encourage stakeholders to behave in a manner more consistent with the regulatory objective. From a psychological standpoint, this is because providing choice involves those subject to the regulations to participate in the process, thereby creating a more favourable perception of the regulatory scheme.

In the main, the observed behaviour in the choice treatments runs counter to the theoretical prediction that choice of rule would encourage participants to choose lower-risk suppliers. Pairwise comparisons of the "crude" implied approach rates between treatments where there was no choice of rule and where a rule choice was offered indicate that introducing rule choice has tended to raise the implied approach rate of biosecurity risk material by at least two percentage points. This is because subjects in the choice treatments tended to favour higher-risk suppliers.

Table 24 shows the results of panel regressions for pairwise comparisons of the treatments considered. Specifically, the four models presented show the following comparisons.

- Model 1: Subjects who chose Mechanism A in treatment Choice6 with treatment C1 (that is, the CSP-1 rule with a clearance number of 10 and a monitoring fraction of 0.2 with full information and standard treatment costs).
- Model 2: Subjects who chose Mechanism B in treatment Choice6 with treatment C1-5.03 (that is, the CSP-1 rule with a clearance number of 5 and a monitoring fraction of 0.3 with full information and standard treatment costs).
- Model 3: Subjects who chose Mechanism A in treatment Choice12 with treatment C1-4.12 (that is, the CSP-1 rule with a clearance number of 10 and a monitoring fraction of 0.2 with full information and high treatment costs).
- Model 4: Subjects whose chose Mechanism B in treatment Choice12 with treatment C1-5.03.12 (that is, the CSP-1 rule with a clearance number of 5 and a monitoring fraction of 0.3 with full information and high treatment costs).

The four models largely confirm the findings based on a partial analysis of the data – giving participants the choice or rule tends to result in supplier choices consistent with a higher average approach rate of biosecurity risk material. In all four models, the coefficients for the rule choice dummy variables are positive and larger for the participants who selected into Mechanism B (models 2 and 4). The impact is marginally significant in model 4, but otherwise statistically insignificant.[84] Note that

[84] The respective p-values and 95 per cent confidence intervals associated with the "giving choice dummy" in the models of Table 24 are:
- model 1 – p-value of 0.799 and 95 per cent confidence interval of [-0.070,0.091];
- model 2 – p-value of 0.277 and 95 per cent confidence interval of [-0.022,0.077];
- model 3 – p-value of 0.862 and 95 per cent confidence interval of [-0.080,0.096]; and
- model 4 – p-value of 0.054 and 95 per cent confidence interval of [-0.001,0.076].

the models presented in Table 24 only control for period effects, in part reflecting the small number of subjects in some comparison groups.[85]

Table 24: Treatment difference models reflecting offering choice of inspection rule

Model	(1)	(2)	(3)	(4)
Dependent variable	Implied biosecurity risk material approach rate			
Giving choice dummy	0.010 (0.041)	0.028 (0.025)	0.008 (0.045)	0.038* (0.020)
Period time trend	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	-0.000 (0.000)
Constant	0.180*** (0.021)	0.207*** (0.020)	0.156*** (0.042)	0.182*** (0.015)
Number of observed choices	1350	1450	550	1700
Subjects	27	29	11	34
Comparison treatment	C1	C1-5.03	C1-4.12	C1-5.03.12

Notes: *, ** and *** indicate that the relevant coefficient is statistically different from zero at the 10, 5 and 1 per cent level of significance respectively. Figures in parentheses refer to standard errors of the estimated model coefficients. Models estimated using feasible generalised least squares with the "standard" random-effects estimator.

[85] The results from adding individual-level controls into these models suggested they would be over-specified. Estimated standard errors were significantly inflated and there were other signs suggesting problems with multicollinearity from models with individual-level controls.

6. The Role of Individual Characteristics in Choosing Suppliers

In this chapter, we investigate the role played by individual characteristics on experimental subjects' supplier choices and the implications these have for the approach rate of biosecurity risk material in consignments. We focus on five main characteristics: risk preferences; environmental concerns; political attitudes; subjects' understanding of the rules; and demographic characteristics, such as nationality. Other than factors associated with understanding the inspection rules, these types of characteristics are not under the control of the biosecurity regulator; rather, they are factors that need to be controlled for in assessing treatment effects in the experimental setting.

Most of the measures assessed in this section come from responses to the post-experiment questionnaire delivered to participants through z-Tree following the main experimental task. Our assessments of individual risk preferences and understanding of the inspection rule also drew upon choices from the first and fourth experimental tasks, respectively. Subjects' performance on the paper-based task and a self-reported measure of understanding the rule from the questionnaire are both used as possible ways to assess the influence of understanding the inspection rule on the choices made in the biosecurity inspection task.

To estimate the role of each of these individual characteristics, we estimate linear regression models,[86] where the dependent variable is the approach rate of biosecurity risk material associated with each supplier choice. For explanatory variables in these statistical models, we use the individual measures of these characteristics and control for treatment and period effects. The results of these regressions are included in Table 25. Each column represents a different statistical model including the individual measures one at a time, with the last column (numbered 11) including all the characteristics together as a robustness check. We interpret these findings in the following sections.

6.1 Risk preferences

The choices in the main experiment involve risk-return trade-offs. Specifically, choosing a supplier with a higher approach rate for biosecurity risk material entails a greater probability of failing an inspection, and thereby achieving a lower financial return, than choosing a supplier with a lower approach rate. However, the "upside" potential if a consignment is not inspected is higher for the experimental subject if they choose a higher-risk (lower-cost) supplier. Therefore, we expect individual preferences for risk-taking will have some influence on the supplier choices we observe in the experimental setting. Consequently, we hypothesise that individuals who are more willing to take risks are more likely to choose supplier with higher probabilities of biosecurity risk material.

[86] As done in Chapter 5, these were estimated as random effects panel-data models using feasible generalised least squares.

Table 25: Statistical models to assess the role of individual characteristics in supplier choices and approach rates of biosecurity risk material

Model	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Dependent variable	Implied biosecurity risk material approach rate										
Lottery choice (Task 1)	0.003 (0.003)										0.001 (0.003)
General risk preference (questionnaire)		0.006*** (0.002)									0.006*** (0.002)
Relative importance of environment/economy			-0.002 (0.003)								-0.007** (0.003)
Concerns for pests and diseases				0.004 (0.005)							0.005 (0.005)
Extinction concerns					0.003 (0.004)						0.006 (0.005)
Belief in successful government intervention						0.002 (0.005)					0.004 (0.005)
Level of conservatism							0.001 (0.002)				0.001 (0.002)
Rule understanding (Task 4)								-0.006 (0.005)			0.003 (0.005)
Rule understanding (questionnaire)									-0.003** (0.001)		-0.002* (0.001)
Non-Australian										0.045*** (0.010)	0.047*** (0.010)
Male											-0.008 (0.009)

Notes: *, ** and *** indicate that the relevant coefficient is statistically different from zero at the 10, 5 and 1 per cent level of significance respectively. Figures in parentheses refer to standard errors of the estimated model coefficients. Estimates of the coefficients for the treatment and period controls and the constant term have been suppressed. All regression equations have 13 750 observations from 50 observations for each of the 275 experimental subjects. Models estimated via feasible generalised least squares using the "standard" random-effects estimator.

We use the lottery-choice task – the first task subjects completed – in addition to the self-reported measure of willingness to take on risks from the post-experiment questionnaire as two potential measures this characteristic. Both of these measures are frequently used in the experimental economics literature to assess risk preferences.

Figure 18 illustrates the average approach rate for biosecurity risk material based on the supplier choices made by subjects in each period for two groups of participants – those who were more risk-averse and those who were more risk-taking. We generated these groups from the responses in the questionnaire on the general willingness to take risks by using a median split. This meant that those who scored that survey question on the 11-point scale between 1 and 6 were classified as "risk-avoiders", while those who scored between 7 and 11 were classified as "risk-takers".

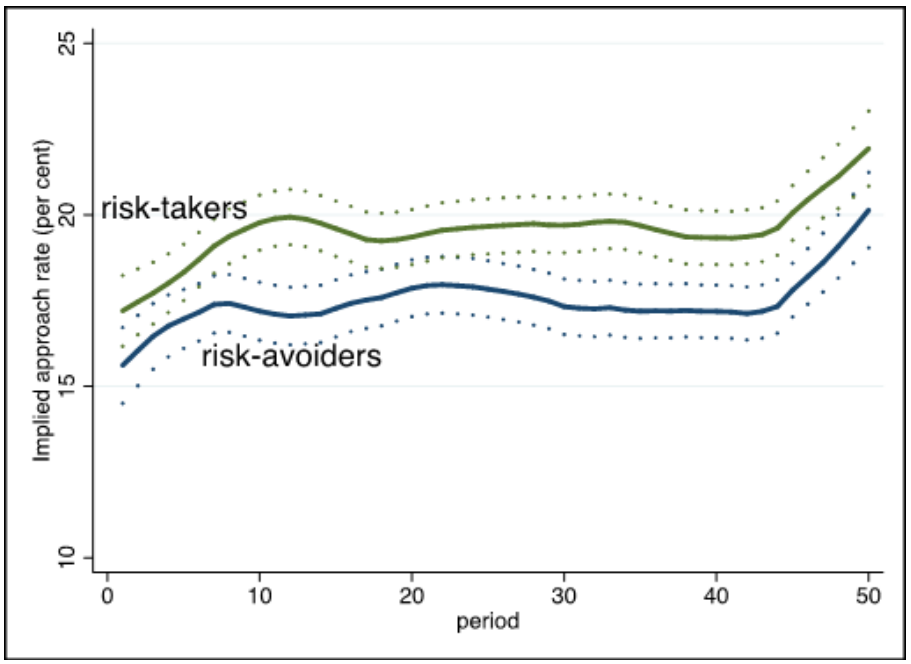


Figure 18: Average approach rate for biosecurity risk material based on supplier choices, accounting for individuals' willingness to take general risks

Notes: Willingness to take general risks is a self-reported measure from the post-experimental questionnaire. A median split of the questionnaire response is used to categorise "risk-takers" and "risk-avoiders".

We observe that risk preferences play an important role, with risk-averse individuals tending to choose suppliers with lower approach rates than those who were more willing to take risks. Models 2 and 11 in Table 25 show that the response to this risk question in the survey is an important predictor for the approach-rate characterisation of supplier choices throughout the experiment. For both models, the coefficient on the dummy variable for being a "risk-taker" is positive and statistically different from zero at the 1 per cent level of significance.[87]

[87] In both models 2 and 11, the p-values associated with the test of the coefficient attached to the "General risk preference (questionnaire)" variable being zero were 0.001 (correct to three decimal places). The respective 95 per cent confidence interval estimates for these coefficients were [0.003,0.010] for both models 2 and 11, when the interval bounds are expressed correct to three decimal places.

We also investigated whether the subjects' choice of gamble in the lottery-choice task provided similar confirmatory evidence. While the choice of gamble in the first experimental task is positively associated with biosecurity risk, as suggested by the positive coefficients for the "Lottery choice (Task 1)" row for models 1 and 11 in Table 25. However, these coefficients are not statistically significant from zero at the usual levels of significance.[88]

Overall, the experimental results provide evidence supporting the notion that individual risk preferences affect supplier choices, which in turn affect the approach rate of biosecurity risk material, in the experiments.

6.2 Attitudes potentially elicited by the experiment's natural frame

6.2.1 Attitudes to the environment

The experiment is embedded in a natural frame and the participants' choices involve selecting suppliers with different approach rates for biosecurity risk material. While there is no actual impact on Australia's biosecurity status as part of the choices made in the experiment, it seems plausible that individuals who are more concerned about the environment may be more likely to avoid suppliers with higher approach rates.

We collected three different proxies for environmental concerns in the post-experimental survey, namely:

- perceptions of the relative importance of the environment compared to the economy;
- concerns about introducing pests and diseases; and
- concerns about the extinction of endangered plants and animals.

In Table 25, the models in the columns labelled 3 to 5 indicate that none of these three variables, on a partial correlation basis, is significantly related to the supplier choices in the main experimental task.[89]

In model 11 of this table, which controls for multiple individual characteristics in the one statistical model, we find that those individuals who rank the environment as more important than the economy tend to choose suppliers with lower approach rates. However, this finding may not be robust, given that the coefficients on the other two environmental concerns measures are insignificant and have positive signs in model 11.[90] A plausible explanation for these results that the partial effect of the

[88] The p-values associated with the tests of the "Lottery choice (Task 1)" coefficients in models 1 and 11 are 0.290 and 0.618, respectively. The respective 95 per cent confidence interval estimates for these coefficients were [-0.002,0.008] for model 1 and [-0.004,0.007] for model 11.

[89] The p-values and interval estimates associated the coefficients of interest in models 3, 4 and 5 are:
- relative importance of the environment – p-value of 0.506 and 95 per cent confidence interval of [-0.008,0.004];
- introducing pests and diseases – p-value of 0.372 and 95 per cent confidence interval of [-0.005,0.013]; and
- extinction concerns – p-value of 0.528 and 95 per cent confidence interval of [-0.006,0.011].

[90] The p-values and interval estimates associated coefficients reflecting attitudes to the environment are:
- relative importance of the environment – p-value of 0.037 and 95 per cent confidence interval of [-0.014,0.000], correct to three decimal places;

"environment versus economy" indicator is accounting for the behaviour of a few individuals in the experiment for which this measure provides some "differentiation" relative to others.

Overall, the results suggest that the framing of the main task around potential environmental concerns has had little impact on subjects' supplier choices.

6.2.2 Political attitudes

In addition to risk preferences and environmental concerns, we measured participants' political attitudes in the post-experiment questionnaire. More precisely, we asked them how well they think governments solve environmental problems and how liberal or conservative in their political leanings they consider themselves. Models 6, 7 and 11 in Table 25 suggest than neither measure of political attitudes was strongly associated with subjects' supplier choices.[91] This finding supports the proposition that the natural framing of the experiment had little impact on the overall results.

6.3 Understanding of the inspection rules

The inspection rules governing the frequency of biosecurity inspections are non-trivial in most of the experimental treatments. This is particularly the case for the three treatments based on the CSP-3 algorithm – a relatively more complicated rule than the CSP-1 algorithm used in 13 treatments. It seems likely many participants could struggle to understand the inspection rules and this, in turn, may affect the supplier choices they make in the main experimental task. Our analysis used the two measures of rule understanding – the self-reported measure and the more objective paper-based incentivised task results – to assess how this factor could affect supplier choices overall.

Models 9 and 11 in Table 25 indicate there is fairly robust evidence that participants who report to understand the rules better tend to choose suppliers associated with lower approach rates of biosecurity risk material.[92] Given the overall tendency of participants to make "riskier" supplier choices compared to the risk-neutral theoretical predictions, the results suggest that a participant who *perceives* to comprehend the

- introducing pests and diseases – p-value of 0.356 and 95 per cent confidence interval of [-0.005,0.015]; and
- extinction concerns – p-value of 0.243 and 95 per cent confidence interval of [-0.004,0.016].

[91] For the indicator assessing perceptions of government intervention to solve environmental problems, the p-values and interval estimates associated with the coefficients in the relevant models are:

- model 6 – p-value of 0.642 and 95 per cent confidence interval of [-0.007,0.012]; and
- model 11 – p-value of 0.416 and 95 per cent confidence interval of [-0.006,0.013].

For the measure of political leanings, the p-values for the coefficients in models 7 and 11 are 0.555 and 0.543 respectively and the 95 per cent confidence interval estimates are [-0.003,0.005] for both models 7 and 11, correct to three decimal places.

[92] For the indicator of understanding based on the questionnaire response, the p-values and interval estimates associated with the significance of the relevant coefficients is:

- model 9 – p-value of 0.034 and 95 per cent confidence interval of [-0.006,0.000], correct to three decimal places; and
- model 11 – p-value of 0.079 and 95 per cent confidence interval of [-0.005,0.000], again correct to three decimal places.

rule better will make choices more consistent with the theoretical predictions and the biosecurity regulator's primary objective.

For the rule test measure arising from the paper-based task, there does not appear to be a statistically significant relationship between that measure of rule understanding and supplier choice. The coefficients on the variable "Rule understanding (Task 4)" in models 8 and 11 of Table 25 are not significantly different from zero.[93] Furthermore, the direction of the relationship does not seem particularly robust, given the sign of the coefficient changes when adding additional covariates in model 11.

6.4 Demographic factors

We also investigated the role of gender and nationality for supplier choice. In Chapter 4, we showed that the subject pool for these experiments was broadly balanced with respect to gender, while more than 70 per cent of the subjects came from Asian countries. While the demographic characteristics of importers, including their nationality, are not within the control of the department, our experimental results could highlight norms that could affect underlying attitudes towards biosecurity compliance in Australia.[94] Given the natural framing of the experimental task which explicitly mentions importing products into Australia, this may be particularly pertinent when comparing domestic students with those from other countries.

Model 11 in Table 25 suggests that gender does not influence supplier choice in aggregate;[95] however, we find a strong impact of nationality on supplier choice,[96] consistent with several of the panel-data regression models shown in Chapter 5.[97] More precisely, Figure 19 shows that Australian experimental subjects are more likely to choose suppliers with lower approach rates of biosecurity risk material than are

[93] The respective p-values and interval estimates attached to the coefficients relating to the Task 4 scores are:
- model 8 – p-value of 0.195 and 95 per cent confidence interval of [-0.015,0.003]; and
- model 11 – p-value of 0.540 and 95 per cent confidence interval of [-0.006,0.012].

[94] At a practical level, this might indicate that a different communication strategy may help encourage improved compliance from all importers. This could be done through an alternative framing of advice around biosecurity compliance which focuses on the private business costs from non-compliance rather than solely appealing to benefits of sustaining Australia's high plant and animal health status and unique environment. We note, however, that this type of examination was beyond the scope of the current experiment, but is a useful conjecture that could be assessed in a future experiment.

[95] The p-value for testing the coefficient on the male indicator in model 11 is different from zero is 0.370, with an associated 95 per cent confidence interval of [-0.026,0.010].

[96] This difference in supplier choice cannot be attributed to the distribution of Australian subjects between the various treatments investigated. On average, Australian participants were slightly over-represented in treatments with higher implied approach rates relative to their international counterparts. While treatment-by-treatment comparisons of the effect of nationality would be interesting to investigate, meaningful comparisons are hampered by the relatively small number of Australians in each treatment group. Most treatments include five or fewer Australian nationals, with the group for treatment C1-I containing no Australians at all.

[97] Treatment comparisons demonstrating sizeable positive (and statistically significant) effects on the implied approach rate of not being an Australian national were for different inspection rules under full or incomplete information (Table 14) and the gain and loss frames for rule feedback under full information (Tables 16 and 17).

their international student counterparts.[98] Models 10 and 11 in Table 25 show the statistically and economically significant positive coefficient for the non-Australian dummy variable, with p-values associated with the tests of these coefficients being less than 0.0005 (zero correct to three decimal places) in both cases.[99]

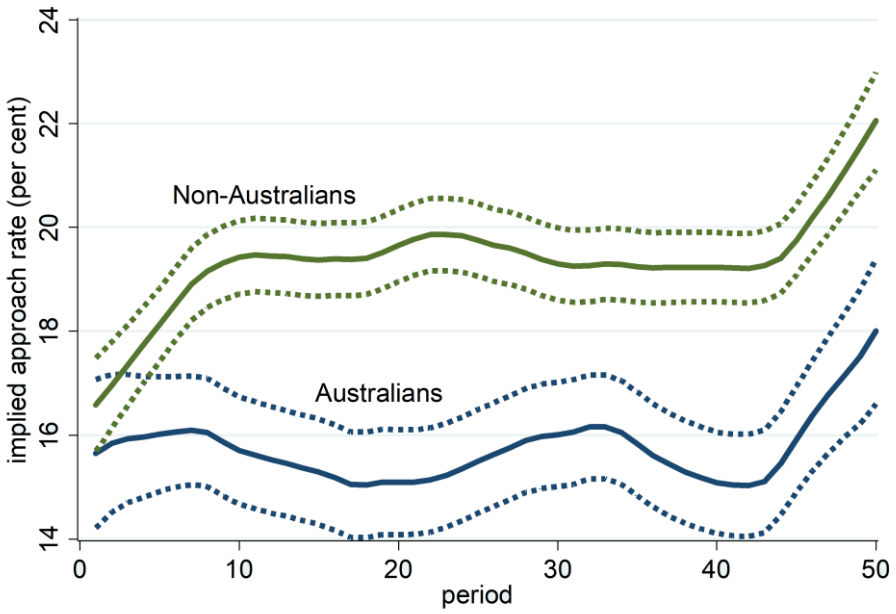


Figure 19: Average approach rate for biosecurity risk material based on supplier choices, accounting for nationality of experimental subjects

While there may be several drivers behind the nationality finding,[100] it is unlikely that subjects' understanding of the rule or environmental or political perceptions are responsible. This is because we control for these individual measures in model 11 and the non-Australian dummy variable's coefficient changes only marginally.

98 The magnitude of differences seemed to be similarly large for students from Asian countries (who constituted a significant share of the sample, as already noted) or those from Europe, Africa, Oceania or the Americas.

99 The 95 per cent confidence intervals associated with the non-Australian dummy variable in the relevant models from Table 25 are [0.025,0.065] for model 10 and [0.026,0.068] for model 11.

100 One plausible explanation is that this reflects a social norm of Australians being aware of the uniqueness of their flora and fauna and of seeking to protect the environment. Participants may have also been more aware of biosecurity issues, given state-based quarantine measures exist around Australia and are frequently seen at major domestic airports. These could be conjectured as explanations, but cannot be verified based on the information collected in the post-experiment questionnaire, since there were no questions in that questionnaire that would be explicitly linked to measuring social or societal norms.

7. Bibliography

Anderson, L.R. and Mellor, J.M. 2009, "Are risk preferences stable? Comparing an experimental measure with a validated survey-based measure", *Journal of Risk and Uncertainty*, 39(2), 137-160.

Avenhaus, R., von Stengel, B. and Zamir, S. 2002, "Inspection Games", in Aumann, R. and Hart, S. (eds), *Handbook of Game Theory with Economic Applications: Volume 3*, North-Holland, Amsterdam, 1947-1987.

Beale, R., Fairbrother, J., Inglis, A. and Trebeck, D. 2008, *One Biosecurity: a working partnership*, Commonwealth of Australia, Canberra, 30 September.

Becker, G.M., DeGroot, M.H. and Marschak, J. 1964, "Measuring utility by a single-response sequential method", *Behavioral Science*, 9(3), 226-232.

Bolton, G.E. 2013, "Not up to Standard: Stress Testing Market Designs for Misbehavior", in Vulkan, N., Roth, A.E. and Neeman, Z. (eds), *The Handbook of Market Design*, Oxford University Press, Oxford, 51-61.

Brewer, P.J. and Plott, C.R. 2002, "A Decentralized, Smart Market Solution to a Class of Back-Haul Transportation Problems: Concept and Experimental Test Beds", *Interfaces*, 32(5), 13-36.

Camerer, C.F. and Hogarth, R.M. 1999, "The Effects of Financial Incentives in Experiments: A Review and Capital-Labor-Production Framework", *Journal of Risk and Uncertainty*, 19(1), 7-42.

Cason, T.N. and Plott, C.R. 1996, "EPA's New Emissions Trading Mechanism: A Laboratory Evaluation", *Journal of Environmental Economics and Management*, 30(2), 133-160.

Charness, G. and Fehr, E. 2015, "From the lab to the real world", *Science*, 350(6020), 512-513.

Charness, G., Gneezy, U. and Imas, A. 2013, "Experimental methods: Eliciting risk preferences", *Journal of Economic Behavior and Organization*, 87, 43-51.

Chen, K.-Y. 2005, "An Economics Wind Tunnel: The Science of Business Engineering", in Morgan, J. (ed), *Advances in Applied Microeconomics: Experimental and Behavioural Economics*, Volume 13, Elsevier, Amsterdam, 133-167.

Cleave, B.L., Nikiforakis, N. and Slonim, R. 2013, "Is there selection bias in laboratory experiments? The case of social and risk preferences", *Experimental Economics*, 16(3), 372-382.

Cox, J.C. and Isaac, R.M. 1987, "Mechanisms for Incentive Regulation: Theory and Experiment", *RAND Journal of Economics*, 18(3), 348-359.

Dave, C., Eckel, C., Johnson, C.A. and Rojas, C. 2010, "Eliciting risk preferences: When is simple better?", *Journal of Rik and Uncertainty*, 41(3), 219-243.

Dodge, H.F. 1943, "A Sampling Inspection Plan for Continuous Production", *The Annals of Mathematical Statistics*, 14(3), 264-279.

Dodge, H.F. and Torrey, M.N. 1951, "Additional continuous sampling inspection regimes", *Industrial Quality Control*, 7(5), 7-12.

Dohmen, T., Falk, A., Huffman, D., Sunde, U., Schupp, J. and Wagner, G.G. 2011, "Individual Risk Attitudes: Measurement, Determinants, and Behavioral Consequences", *Journal of the European Economic Association*, 9(3), 522-550.

Duffy, J. and Fehr, D. 2017, "Equilibrium selection in similar repeated games: experimental evidence on the role of precedents", forthcoming in *Experimental Economics*. Online version available at: https://link.springer.com/article/10.1007/s10683-017-9531-6.

Dulleck, U., Fooken, J. and Fell, J. 2013, "Within-Subject Intra- and Inter-Method Consistency of Two Experimental Risk Attitude Elicitation Methods", *German Economic Review*, 16(1), 104-121.

Eckel, C. and Grossman, P. 2002, "Sex differences and statistical stereotyping in attitudes toward financial risk", *Evolution and Human Behavior*, 23 (4), 281-295.

Eckel, C. and Grossman, P. 2008, "Forecasting risk attitudes: An experimental study using actual and forecast gamble choices", *Journal of Economic Behavior and Organization*, 68(1), 1-17.

Elliott, C.S. and Hayward, D.M. 1998, "The Expanding Definition of Framing and Its Particular Impact on Economic Experimentation", *Journal of Socio-Economics*, 27(2), 229-243.

Falk, A. and Heckman, J.J. 2009, "Lab Experiments Are a Major Source of Knowledge in the Social Sciences", *Science*, 326 (5952), 535-538.

Fischbacher, U. 2007, "z-Tree: Zurich toolbox for ready-made economic experiments", *Experimental Economics*, 10(2), 171-178.

Frechette, G.R. and Schotter, A. (eds) 2015, *Handbook of Experimental Economic Methodology*, Oxford University Press, New York

Friedman, F. and Cassar, A. 2004, *Economics lab: An intensive course in experimental economics*, Routledge, London.

Goeree, J.K., Holt, C.A. and Ledyard, J.O. 2006, "An Experimental Comparison of the FCC's Combinatorial and Non-Combinatorial Simultaneous Multiple Round Auctions", Prepared for the Wireless Telecommunications Bureau of the Federal Communications Commission, July 12. Available at: wireless.fcc.gov/auctions/data/papersAndStudies/fcc_final_report_071206.pdf.

Greenberg, J. 1984, "Avoiding Tax Avoidance: A (Repeated) Game-Theoretic Approach", *Journal of Economic Theory*, 32(1), 1-13.

Greene, W.H. 2003, *Econometric Analysis*, Fifth Edition, Pearson Education, Upper Saddle River.

Greiner, B. 2015, "Subject pool recruitment procedures: organizing experiments with ORSEE", *Journal of the Economic Science Association*, 1(1), 114-125.

Grether, D.M., Isaac, R.M. and Plott, C.R. 1981, "The Allocation of Landing Rights by Unanimity Among Competitors", *American Economic Review: Papers and Proceedings*, 71(2), 166-171.

Grether, D.M. and Plott, C.R. 1984, "The Effects of Market Practices in Oligopolistic Markets: An Experimental Examination of the Ethyl Case", *Economic Inquiry*, 22(4), 479-507.

Guala, F. 2005, *The Methodology of Experimental Economics*, Cambridge University Press, New York.

Harrison, G.W., Lau, M.I. and Rütstrom, E.E. 2009, "Risk attitudes, randomization to treatment, and self-selection into experiments", *Journal of Economic Behavior and Organization*, 70(3), 498–507.

Harrison, G.W. and List, J.A. 2004, "Field Experiments", *Journal of Economic Literature*, 42(4), 1009-1055.

Harrison, G.W. and Rütstrom, E.E. 2008, "Risk Aversion in the Laboratory", in Cox, J.C. and Harrison, G.W. (eds), *Risk Aversion in Experiments*, Research Methods in Experimental Economics, Volume 12, 41-196.

Harrison, G.W. and Rütstrom, E.E. 2009, "Expected utility theory and prospect theory: one wedding and a decent funeral", *Experimental Economics*, 12(2), 133-158.

Herbst, D. and Mas, A. 2015, "Peer effects on worker output in the laboratory generalize to the field", *Science*, 350(6260), 545-549.

Hey, J.D., Morone, A. and Schmidt, U. 2009, "Noise and bias in eliciting preferences", *Journal of Risk and Uncertainty*, 39(3), 213-235.

Hertwig, R. and Ortmann, A. 2001, "Experimental practices in economics: A methodological challenge for psychologists?", *Behavioral and Brain Sciences*, 24(3), 383-403.

Holt, C.A. and Laury, S.K. 2002, "Risk Aversion and Incentive Effects", *American Economic Review*, 92(5), 1644-1655.

Holt, C.A. and Laury, S.K. 2005, "Risk Aversion and Incentive Effects: New Data Without Order Effects", *American Economic Review*, 95(3), 902-904.

Holt, C.A. and Laury, S.K. 2014, "Assessment and Estimation of Risk Preferences", in Machina, M.J. and Viscusi, W.K. (eds), *Handbook of the Economics of Risk and Uncertainty: Volume 1*, North-Holland, Oxford, 135-201.

Hsiao, C. 2014, *Analysis of Panel Data*, Third Edition, Cambridge University Press, New York.

Jehle, G.A. and Reny, P.J. 2001, *Advanced Microeconomic Theory*, Second Edition, Addison Wesley, Boston.

Jiminez-Buedo, M. and Miller, L.M. 2010, "Why a Trade-Off? The Relationship between the External and Internal Validity of Experiments", *Theoria: An International Journal for Theory, History and Foundations of Science*, 25(3), 301-321.

Kessler, J.B. and Vesterlund, L. 2015, "The External Validity of Laboratory Experiments: The Misleading Emphasis on Quantitative Effects", in Frechette, G.R. and Schotter, A. (eds), *Handbook of Experimental Economic Methodology*, Oxford University Press, New York, 391-406.

Koudstaal, M., Sloof, R. and van Praag, M. 2016, "Risk, Uncertainty, and Entrepreneurship: Evidence from a Lab-in-the-Field Experiment", *Management Science*, 62(10), 2897-2915.

Laffont, J.-J. and Tirole, J. 1993, *A Theory of Incentives in Procurement and Regulation*, MIT Press, Cambridge: MA.

Levitt, S.D. and List, J.A. 2009, "Field experiments in economics: The past, the present, and the future", *European Economic Review*, 53(1), 1-18.

List, J.A. 2008, "Informed Consent in Social Science: Response", *Science*, 322(5902), 672.

List, J., Shaikh, A.M. and Xu, Y. 2016, "Multiple Hypothesis Testing in Experimental Economics", unpublished manuscript, 22 November. Available at http://home.uchicago.edu/amshaikh/webfiles/experimental.pdf.

Loewenstein, G. 1999, "Experimental Economics from the Vantage-point of Behavioural Economics", *Economic Journal*, 109(453), F25-F34.

Lönnqvist, J.-E., Verkasalo, M., Walkowitz, G. and Wichardt, P.C. 2015, "Measuring individual risk attitudes in the lab: Task or ask? An empirical comparison", *Journal of Economic Behavior and Organization*, 119, 254-266.

Lugovskyy, V., Puzzello, D., Sorensen, A., Walker, J. and Williams, A. 2017, "An experimental study of finitely and infinitely repeated linear public goods games", *Games and Economic Behavior*, 102, 286-302.

National Health and Medical Research Council, Australian Research Council and the Australian Vice-Chancellors' Committee 2015, *National Statement on Ethical Conduct in Human Research 2007 (Updated May 2015)*, Commonwealth of Australia, Canberra. Available at: http://www.nhmrc.gov.au/_files_nhmrc/publications/attachments/e72_national_statement_may_2015_150514_a.pdf.

Nemes, V., Plott, C.R. and Stoneham, G.S. 2008, "Electronic BushBroker exchange: Designing a combinatorial double auction for native vegetation offsets", unpublished working paper, July. Available at: http://ssrn.com/abstract=1212202.

Plott, C.R. 1987, "Dimensions of Parallelism: Some Policy Applications of Experimental Methods", in Roth, A.E. (ed), *Laboratory Experimentation in Economics: Six Points of View*, Cambridge University Press, New York, 193-219.

Plott, C.R. 1997, "Laboratory Testing Testbeds: Application to the PCS Auction", *Journal of Economics and Management Strategy*, 6(3), 605–638.

Robinson, A., Bell, J., Woolcott, B. and Perotti, E. 2012, *AQIS Quarantine Operations Risk Return ACERA 1001 Study J Imported Plant-Product Pathways*, Australian Centre of Excellence for Risk Analysis, University of Melbourne, Project 1001J, Final Report. Available at: http://www.acera.unimelb.edu.au/materials/endorsed/1001j.pdf.

Roe, B.E. and Just, D.R. 2009, "Internal and External Validity in Economics Research: Tradeoffs between Experiments, Field Experiments, Natural Experiments, and Field Data", *American Journal of Agricultural Economics*, 91(5), 1266-1271.

Rossiter, A. and Hester, S. 2017, "Designing biosecurity inspection regimes to account for stakeholder incentives: An inspection game approach", *Economic Record*, 93(301), 277-301.

Rossiter, A., Hester, S., Aston, C., Sibley, J., Stoneham, G. and Woodhams, F. 2016, *Incentives for Importer Choices*. Centre of Excellence for Biosecurity Risk Analysis, University of Melbourne, Project 1304C, Final Report 1: Overview, 14 November. Available at: http://cebra.unimelb.edu.au/__data/assets/pdf_file/0020/2172152/CEBRA-Project-1304C-Final-Report.pdf.

Roth, A.E. 1995, "Introduction to Experimental Economics", in Hagel, J.H. and Roth, A.E. (eds), *The Handbook of Experimental Economics*, Princeton University Press, Princeton, 3-109.

Roth, A.E. 2002, "The Economist as Engineer: Game Theory, Experimentation, and Computation as Tools for Design Economics", *Econometrica*, 70(4), 1341–1378.

Roth, A.E. 2013, "What Have We Learned From Market Design?", in Vulkan, N., Roth, A.E. and Neeman, Z. (eds), *The Handbook of Market Design*, Oxford University Press, Oxford, 7-50.

Roth, A.E. and Murnighan, J.K. 1978, "Equilibrium Behavior and Repeated Play of the Prisoner's Dilemma", *Journal of Mathematical Psychology*, 17(2), 189-198.

Sappington, D. 1994, "Designing Incentive Regulation", *Review of Industrial Organization*, 9(3), 245-272.

Smith, V.L.1976, "Experimental Economics: Induced Value Theory", *American Economic Review: Papers and Proceedings*, 66(2), 274-279.

Smith, V.L. 1994, "Economics in the Laboratory", *Journal of Economic Perspectives*, 8(1), 113-131.

Smith, V.L. 2008, "Experimental Economics", in Durlauf, S.N. and Blume, L.E. (eds), *The New Palgrave Dictionary of Economics Online*, Second Edition, Palgrave Macmillan. Available at: http://www.dictionaryofeconomics.com/article?id=pde2008_E000186.

Tversky, A. and Kahneman, D. 1991, "Loss Aversion in Riskless Choice: A Reference-Dependent Model", *Quarterly Journal of Economics*, 106(4), 1039-1061.

Weber. E.U. 2013, "Doing the Right Thing Willingly", in Shafir, E. (ed), *The Behavioral Foundations of Public Policy*, Princeton University Press, Princeton, 380-397.

8. List of Figures

9. List of Tables

Appendix A: Experimental Procedures and Instructions

A.1. Computer-based experimental task instructions

A.1.1. Preamble and Task 1: Eckel and Grossman (2002, 2008) risk preference elicitation task

SCREEN 1

WELCOME

By participating in this experiment, you and all other participants receive an endowment of $5.

In addition, there is the possibility to increase your earnings, which will be paid to you in private after the experiment has finished.

We will now explain how.

Please click ‘continue’ on your computer screen.

SCREEN 2

Two tasks

Today’s experiment consists of two tasks.

You can increase your earnings in both tasks.

On the next screen you will see the instructions for the first task.

Please click ‘continue’ on your computer screen.

SCREEN 3

Task 1

In the first task, you will only make one choice. Your choice is to select one gamble from five different gambles.

For every gamble, each event (A and B) has a 50% chance of occurring. The computer software will randomly determine whether event A or B occurs. Here are the five gambles:

Gamble 1:	Event A:	$4.50 with 50% probability
	Event B:	$3.50 with 50% probability
Gamble 2:	Event A:	$6 with 50% probability
	Event B:	$3 with 50% probability
Gamble 3:	Event A:	$8 with 50% probability
	Event B:	$2 with 50% probability
Gamble 4:	Event A:	$10 with 50% probability
	Event B:	$1 with 50% probability
Gamble 5:	Event A:	$12 with 50% probability
	Event B:	$0 with 50% probability

Example 1

If you select gamble 3 and the software randomly determines event A, $8 will be added to your earnings. If the software determines event B, $2 will be added to your earnings.

Example 2

If you select gamble 5 and the software randomly determines event A, $12 will be added to your earnings. If the software determines event B, $0 will be added to your earnings.

Which gamble do you want to play? {insert radio buttons 1,2,3,4,5}

If you have any questions, please raise your hand and we will come to you.

Please press ‘continue’ now on your computer after you have made your choice.

SCREEN 4

You have now finished Task 1.

We will provide your outcome in Task 1 at the end of today's experiment.

Before that, you will do Task 2.

Please click 'continue' on your computer screen to proceed to Task 2, the last task in this experiment.

A.1.2. Task 2: Biosecurity inspection game task (screens common to all treatments)

SCREEN 5

Task 2

In the second task, you will make 50 choices in the position of an importer.

As an importer, you will buy shipments containing ten goods to bring to Australia.

Your choice is always to select one supplier from four different options for each shipment.

Each of your 50 shipments is worth 200 monetary units (MU); i.e. each good is worth 20 MU.

At the end of the experiment we will convert the monetary units into Australian dollars at the exchange rate:

400 MU = $1

Please press 'continue' now on your computer and we will hand out detailed information on each of the four supplier options.

If you have any questions, please raise your hand and we will come to you.

SCREEN 6

Please read your handout and keep it at your place.

On the following screen, we will show some examples.

Please press 'continue' now on your computer if you have no questions about your information sheet.

If you have any questions, please raise your hand and we will come to you.

SCREEN 7

This is treatment-dependent and the information on this screen repeats the information contained in the handout described in Appendix A.3 below.

SCREEN 8

Practice Trial

Before you start with your first of the 50 shipments, there is a practice trial to make you are familiar with the decision environment. The outcomes in the trial do not affect your earnings.

Please press continue to start the trial.

SCREEN 9

Start of Practice Trial

TRIAL

Which supplier option do you choose for your shipment trial?

Supplier option: A, B, C, D {radio buttons}

Please press 'continue' now on your computer if you have no questions.

If you have any questions, please raise your hand and we will come to you.

SCREEN 10

TRIAL

Outcomes for shipment trial: *{example; depending on treatment}*[101]

You chose supplier option	**A**									
Value of good	**20**	**20**	**20**	**20**	**20**	**20**	**20**	**20**	**20**	**20**
Transportation and purchase costs	**3**	**3**	**3**	**3**	**3**	**3**	**3**	**3**	**3**	**3**
Was good inspected?	**N**	**N**	**Y**	**Y**	**N**	**N**	**N**	**N**	**N**	**N**
Were biosecurity risks detected?	**-**	**-**	**Y**	**N**	**-**	**-**	**-**	**-**	**-**	**-**
Earnings	**17**	**17**	**-1**	**13**	**17**	**17**	**17**	**17**	**17**	**17**
Total earnings shipment 1 (in MU)	**138**									

SCREEN 11

Please press continue to start with the first shipment of Task 2.

If you have any questions, please raise your hand and we will come to you.

[101] This presentation represents an example of the output that might be expected under treatment R (randomised inspection where each consignment has a 20 per cent probability of inspection).

SCREEN C1.1

SHIPMENT 1 of 50

Which supplier option do you choose for shipment 1?

Supplier option: A, B, C, D {radio buttons}

Please press 'continue' now on your computer if you have no questions.

If you have any questions, please raise your hand and we will come to you.

SCREEN C1.2

SHIPMENT 1 of 50

Outcomes for shipment 1: *{example; depending on treatment}*

You chose supplier option	**A**									
Value of good	**20**	**20**	**20**	**20**	**20**	**20**	**20**	**20**	**20**	**20**
Transportation and purchase costs	**3**	**3**	**3**	**3**	**3**	**3**	**3**	**3**	**3**	**3**
Was good inspected?	**N**	**N**	**Y**	**Y**	**N**	**N**	**N**	**N**	**N**	**N**
Were biosecurity risks detected?	-	-	**Y**	**N**	-	-	-	-	-	-
Earnings	**17**	**17**	**-1**	**13**	**17**	**17**	**17**	**17**	**17**	**17**
Total earnings shipment 1 (in MU)	**138**									

SCREEN C2.1

SHIPMENT 2 of 50

Which supplier option do you choose for shipment 1?

Supplier option: A, B, C, D {radio buttons}

Please press 'continue' now on your computer if you have no questions.

If you have any questions, please raise your hand and we will come to you.

SCREEN C2.2

SHIPMENT 2 of 50

Analogous presentation as for SCREEN C1.2.

Go on until Shipment 50 of 50 (SCREEN C50.2).

SCREEN 12

End of task 2

You have now finished Task 2.

Please press 'continue' to be informed about your earnings.

SCREEN 13

EARNINGS

TASK 1:

You chose gamble {insert} and event {insert} occurred.

Consequently your earnings are {insert}.

TASK 2:

Over the 50 shipments you earned: {insert}

Total earnings: {insert}

Please press 'continue'.

A.1.3. Task 3: Post-experiment questionnaire

These questions all appear on the one screen as part of a different module in z-Tree.

What is your gender?
{radio buttons: female and male}

What is your height?
{text box}

What is your major?
{text box}

What is your nationality?
{text box}

How do you see yourself: are you generally a person who is fully prepared to take risks or do you try to avoid taking risks?
{slider from 1 = not at all willing to take risk, to, 11 = very willing to take risks}

How concerned are you about the risk of introducing exotic pests and diseases to Australia?
{radio buttons: extremely concerned; very concerned; moderately concerned; slightly concerned; not at all concerned}

How concerned are you about the extinction of endangered plants and animals?
{radio buttons: extremely concerned; very concerned; moderately concerned; slightly concerned; not at all concerned}

How important is protecting the environment compared with improving the economy?
{radio buttons: much more important; somewhat more important; slightly more important; about as important; slightly less important; somewhat less important; much less important}

When governments get involved in trying to solve environmental problems, how often do you think they make things better?
{radio buttons: always; most of the time; about half of the time; once in a while; never}

We hear a lot of talk these days about 'liberals' and 'conservatives'. Here is an 11-point scale on which people's political views are arranged from extremely liberal to extremely conservative. Where would you place yourself on this scale?
{slider from 1 = extremely liberal, to, 11 = extremely conservative}

To which kind of school did you go before starting at Monash University?
{radio buttons: private; Catholic; public; international}

Do you have a job to help fund your studies and living expenses?
{radio buttons: no, yes}
If yes: How many hours do you work per week?
{text box}

How well did you understand the mechanism that the quarantine service used to inspect goods?
{slider from 1 = I do not at all understand, to, 11 = completely understand}

A.2. Handout outlining the biosecurity inspection experimental task scenario

A.2.1. Standard handout (13 treatments)

Background information for Task 2

There are 50 shipments.

Each shipment is worth 200 MU.

Each shipment contains 10 goods, thus each good is worth 20 MU.

What do you have to do?

You have to choose one supplier option for each shipment you want to import. There are four supplier options (**A, B, C, D**).

What are differences between the supplier options?

The supplier options differ on two dimensions:

- transport and purchase costs per shipment; and
- the probability that a good in the shipment contains biosecurity risk material. Biosecurity risk material is material that has the potential to introduce an exotic pest or disease into Australia, such as live insects, seeds and soil.

This table illustrates the different supplier options:

Supplier option	**A**	**B**	**C**	**D**
Transportation and purchase costs per shipment (**in MU**)	**30**	**50**	**60**	**70**
Probability that a good in a shipment contains biosecurity risk material***	**50%**	**25%**	**10%**	**2%**

*** Note that a probability of, for example, 50% does not automatically imply that 5 out of the 10 goods in a shipment contain biosecurity risk material but that there is a 50% probability that each single good contains biosecurity risk material. Thus, it is possible that the number of goods in a shipment containing biosecurity risk material is less than, equal to, or greater than 5.

What happens after you have chosen a supplier?

After you have chosen a supplier, your shipment is reviewed by the quarantine service (which is played by the computer software). The quarantine service does not know which supplier option you have selected.

What does quarantine do and how does it affect your earnings?

The quarantine service either inspects or does not inspect a good for biosecurity risk material.

- If a good is not inspected, there will be no costs for you.
- If a good is inspected and no biosecurity risk material is found, there is a cost for you of **4 MU**.
- If a good is inspected and biosecurity risk material is found, there is an **additional** cost for you of **6 MU**. This means that there are total costs of 4 + 6 = 10 MU.

On the next computer screen we will inform you how often inspections take place.

A.2.2. Variations for cost treatments

Across the treatments where costs change, the only section that differs between the treatments is the last section of the handout: **"What does quarantine do and how does it affect your earnings?"** In particular, it is only the second and third dot-points in the list that change as described below.

For Treatment C1-2.6:

- If a good is inspected and no biosecurity risk material is found, there is a cost for you of **2 MU**.
- If a good is inspected and biosecurity risk material is found, there is an **additional** cost for you of **6 MU**. This means that there are total costs of 2 + 6 = 8 MU.

For Treatment C1-2.12:

- If a good is inspected and no biosecurity risk material is found, there is a cost for you of **2 MU**.
- If a good is inspected and biosecurity risk material is found, there is an **additional** cost for you of **12 MU**. This means that there are total costs of 2 + 12 = 14 MU.

For Treatments C1-4.12, C1-5.03.12 and Choice 12:

- If a good is inspected and no biosecurity risk material is found, there is a cost for you of **4 MU**.
- If a good is inspected and biosecurity risk material is found, there is an **additional** cost for you of **12 MU**. This means that there are total costs of 4 + 12 = 16 MU.

A.3. Handouts on the inspection rule for different experimental treatments

A.3.1. Treatment M

How often does quarantine service inspect your goods?

Quarantine service inspects each good with a probability of 100%.

This means that all goods in each shipment are inspected.

Please press 'continue' for a practice trial.

A.3.2. Treatment R

How often does quarantine service inspect your goods?

Quarantine service inspects **each good** with a probability of **20%**.

This means that *on average* 2 out of the 10 goods in each shipment are inspected. However, it is possible that in a given shipment the number of goods actually inspected may be less than, equal to, or greater than 2.

Please press 'continue' for a practice trial.

A.3.3. Treatment C3

How often does quarantine service inspect your goods?

Each of your goods (100%) will be inspected until you meet the requirement for a reduced inspection rate of **20%**.

To meet the requirement for the reduced inspection rate, **you need to pass 10 inspections in a row**. That is, quarantine service needs to find that 10 goods in a row did not contain biosecurity risk material.

At the reduced inspection rate, *on average* 2 out of the 10 goods in each shipment are inspected. However, it is possible that in a given shipment the number of goods actually inspected is less than, equal to, or more than 2.

The reduced inspection rate applies as long as quarantine service does not find a good in your shipment with biosecurity risk material.

If quarantine service finds a good with biosecurity risk material, you lose the reduced inspection rate. You will then return to having all your goods inspected until you pass inspection **4 times in a row** to receive the reduced inspection rate.

However, if you fail an inspection and your last inspection failure was 10 or fewer inspections ago, you will return to having all your goods inspected until you pass inspection **10 times in a row** to receive the reduced inspection rate.

Please press ‘continue’ for a practice trial.

A.3.4. Treatment C3-I

How often does quarantine service inspect your goods?

Each of your goods (100%) will be inspected until you meet the requirement for a reduced inspection rate of **between 10% and 50%**.

To meet the requirement for the reduced inspection rate, **you need to pass 10 inspections in a row**. That is, quarantine service needs to find that 10 goods in a row did not contain biosecurity risk material.

The reduced inspection rate applies as long as quarantine service does not find a good in your shipment with biosecurity risk material.

If quarantine service finds a good with biosecurity risk material, you lose the reduced inspection rate. You will then return to having all your goods inspected until you pass inspection **4 times in a row** to receive the reduced inspection rate.

However, if you fail an inspection and your last inspection failure was 10 or fewer inspections ago, you will return to having all your goods inspected until you pass inspection **10 times in a row** to receive the reduced inspection rate.

Please press ‘continue’ for a practice trial.

A.3.5. Treatment C3-I2

How often does quarantine service inspect your goods?

Each of your goods (100%) will be inspected until you meet the requirement for a reduced inspection rate of **between 10% and 50%**.

To meet the requirement for the reduced inspection rate, **you need to pass 10 inspections in a row**. That is, quarantine service needs to find that 10 goods in a row did not contain biosecurity risk material.

The reduced inspection rate applies as long as quarantine service does not find a good in your shipment with biosecurity risk material.

If quarantine service finds a good with biosecurity risk material, you lose the reduced inspection rate. You will then return to having all your goods inspected until you pass inspection **for the next few goods** to receive the reduced inspection rate.

However, if you fail an inspection and your last inspection failure was 10 or fewer inspections ago, you will return to having all your goods inspected until you pass inspection **10 times in a row** to receive the reduced inspection rate.

Please press ‘continue’ for a practice trial.

A.3.6. Treatments C1, C1-L, C1-G, C1-2.6, C1-2.12 and C1-4.12

How often does quarantine service inspect your goods?

Each of your goods (100%) will be inspected until you meet the requirement for a reduced inspection rate of **20%**.

To meet the requirement for the reduced inspection rate, **you need to pass 10 inspections in a row**. That is, quarantine service needs to find that 10 goods in a row did not contain biosecurity risk material.

At the reduced inspection rate, *on average* 2 out of the 10 goods in each shipment are inspected. However, it is possible that in a given shipment the number of goods actually inspected is less than, equal to, or greater than 2.

The reduced inspection rate applies as long as quarantine service does not find a good in your shipment with biosecurity risk material.

If quarantine service finds a good with biosecurity risk material, you lose the reduced inspection rate. You will then return to having all your goods inspected until you pass inspection **10 times in a row** to receive the reduced inspection rate.

Please press ‘continue’ for a practice trial.

A.3.7. Treatments C1-I, C1-IL and C1-IG

How often does quarantine service inspect your goods?

Each of your goods (100%) will be inspected until you meet the requirement for a reduced inspection rate. The reduced inspection rate is **between 10% and 50%**.

To meet the requirement for the reduced inspection rate, you need to pass 10 inspections in a row. That is, quarantine service needs to find that 10 goods in a row did not contain biosecurity risk material.

The reduced inspection rate applies as long as quarantine service does not find a good in your shipment with biosecurity risk material.

If quarantine service finds a good with biosecurity risk material, you lose the reduced inspection rate. You will then return to having all your goods inspected until you pass inspection **10 times in a row** to receive the reduced inspection rate.

Please press 'continue' for a practice trial.

A.3.8. Treatments C1-5.03 and C1-5.03.12

How often does quarantine service inspect your goods?

Each of your goods (100%) will be inspected until you meet the requirement for a reduced inspection rate of **30%**.

To meet the requirement for the reduced inspection rate, **you need to pass 5 inspections in a row**. That is, quarantine service needs to find that 5 goods in a row did not contain biosecurity risk material.

At the reduced inspection rate, *on average* 3 out of the 10 goods in each shipment are inspected. However, it is possible that in a given shipment the number of goods actually inspected is less than, equal to, or greater than 3.

The reduced inspection rate applies as long as quarantine service does not find a good in your shipment with biosecurity risk material.

If quarantine service finds a good with biosecurity risk material, you lose the reduced inspection rate. You will then return to having all your goods inspected until you pass inspection **5 times in a row** to receive the reduced inspection rate.

Please press 'continue' for a practice trial.

A.3.9. Treatments Choice6 and Choice12

How often does quarantine service inspect your goods?

Quarantine service offers you two different inspection mechanisms and you have to choose one of the two mechanisms for the whole experiment.

Mechanism A:

Each of your goods (100%) will be inspected until you meet the requirement for a reduced inspection rate of **20%**.

To meet the requirement for the reduced inspection rate, **you need to pass 10 inspections in a row**. That is, quarantine service needs to find that 10 goods in a row did not contain biosecurity risk material.

At the reduced inspection rate, *on average* 2 out of the 10 goods in each shipment are inspected. However, it is possible that in a given shipment the number of goods actually inspected is less than, equal to, or greater than 2.

The reduced inspection rate applies as long as quarantine service does not find a good in your shipment with biosecurity risk material.

If quarantine service finds a good with biosecurity risk material, you lose the reduced inspection rate. You will then return to having all your goods inspected until you pass inspection **10 times in a row** to receive the reduced inspection rate.

Mechanism B:

Each of your goods (100%) will be inspected until you meet the requirement for a reduced inspection rate of **30%**.

To meet the requirement for the reduced inspection rate, you need to pass **5 inspections in a row**. That is, quarantine service needs to find that 5 goods in a row did not contain biosecurity risk material.

At the reduced inspection rate, *on average* 3 out of the 10 goods in each shipment are inspected. However, it is possible that in a given shipment the number of goods actually inspected is less than, equal to, or greater than 3.

The reduced inspection rate applies as long as quarantine service does not find a good in your shipment with biosecurity risk material.

If quarantine service finds a good with biosecurity risk material, you lose the reduced inspection rate. You will then return to having all your goods inspected until you pass inspection **5 times in a row** to receive the reduced inspection rate.

Which mechanism do you choose?

{radio button "Mechanism A"; "Mechanism B"}

Please press 'continue' for a practice trial.

A.4. Paper-based task on rule understanding

A.4.1. Standard test format

Additional opportunity to increase your earnings

Please write your computer number here: _____

Please take a look at the two different inspection scenarios for two new importers and fill out the last rows in each scenario. The last rows should correctly identify the probability with which quarantine service inspects a given good in the experiment you just participated in.

To recall how quarantine service inspected your goods during the experiment, take a look at the "How often does quarantine service inspect your goods?" handout.

Write the probability of inspection in each cell. For each correctly filled out row, you will receive **$1**.

Scenario 1: New herbal tea importer

Number of good	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Were biosecurity risks detected?**		Yes												Yes						
Probability of inspection?																				

Scenario 2: New dried vegetable importer

Number of good	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Were biosecurity risks detected?**												Yes	Yes							
Probability of inspection?																				

** To be clear, if biosecurity risk material was detected, that particular good was inspected by the quarantine service. However, the probability of that good being inspected may not be 100%.

A.4.2. Variations from standard test format

Treatment M

This treatment received the CSP-3 algorithm description that treatment C3 received in the second task of the experiment, as well as a diagrammatic representation of the CSP-3 algorithm (see the subsection below). This variation was reflected in the task instructions:

"Please take a look at the two different inspection scenarios for two new importers and fill out the last rows in each scenario. The last rows should correctly identify the probability with which quarantine service inspects a given good. On the next page you will find an explanation about how often quarantine service inspects goods in these two scenarios.

For an illustration on how quarantine service inspects these goods please also see the diagram on the next page."

Treatment R

This treatment received the CSP-3 algorithm description that treatment C3 received in the second task of the experiment, but unlike treatment M received no diagram. This variation was reflected in the task instructions:

"Please take a look at the two different inspection scenarios for two new importers and fill out the last rows in each scenario. The last rows should correctly identify the probability with which quarantine service inspects a given good. On the next page you will find an explanation about how often quarantine service inspects goods in these two scenarios."

Treatments C3 and C3-I

Subjects who received the diagrammatic representation of the CSP-3 algorithm as part of this task had the added line:

"For an illustration on how quarantine service inspects these goods please also see the diagram on the next page."

Treatment C3-I2

To avoid potential confusion in completing this task with the vague description of the tight census parameter, the instructions were modified so it was clear the tight census number was four. In this case, the sentence was added:

"Note that the fourth paragraph should read:

"*If quarantine service finds a good with biosecurity risk material, you lose the reduced inspection rate. You will then return to having all your goods inspected until you pass inspection* ***for the next 4 goods*** *to receive the reduced inspection rate.*""

For those in this treatment receiving the diagram of the CSP-3 algorithm, the line described for treatments C3 and C3-I above was added to those instructions.

Treatments Choice6 and Choice12

To make it clear which inspection rule subjects in these treatments chose, we added a line which asked:

"Which mechanism did you choose (circle): A B"

Given the choice of rule, the second sentence of the task instructions was modified to read:

"The last rows should correctly identify the probability with which the quarantine service inspects a given good in the experiment you just participated in for the mechanism (A or B) you chose."

A.4.3. Diagrammatic representations of the CSP-3 algorithm

Treatments M and C-3

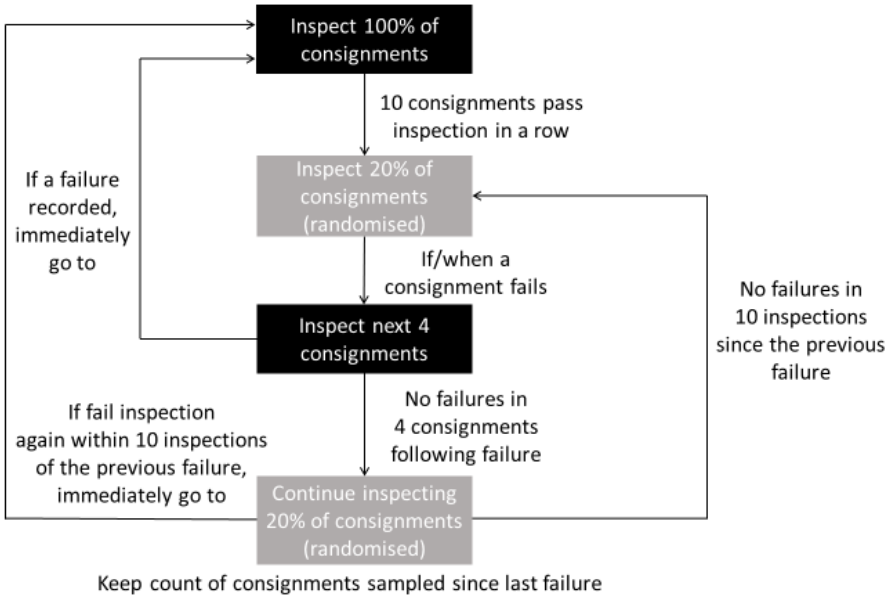


Figure 20: CSP-3 algorithm diagrams for full-information treatments

Treatments C3-I and C3-I2

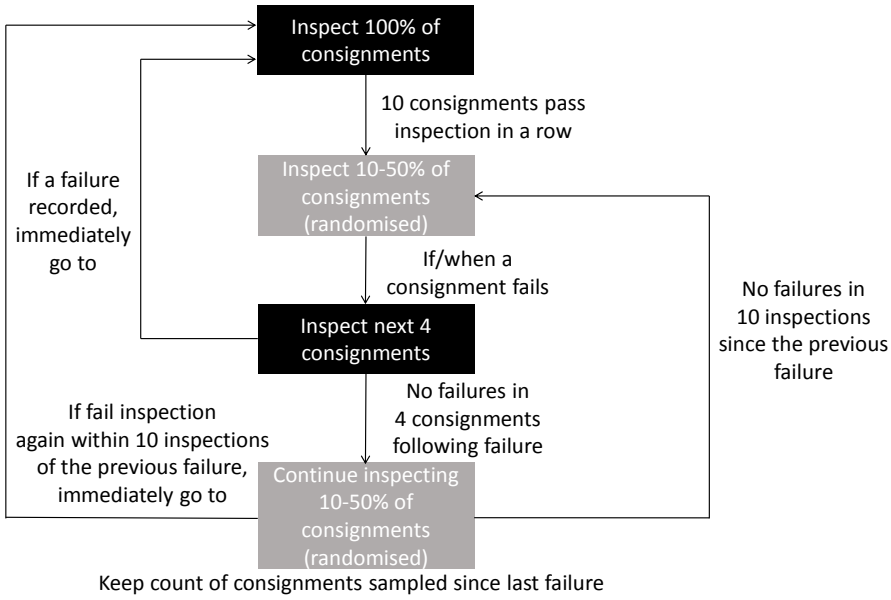


Figure 21: CSP-3 algorithm diagrams for treatments where the monitoring fraction is vaguely described

Appendix B: Simulation Models for Calibrating Parameters

B.1. Overview of simulation strategy

The simulation strategy for the experiment involved choosing a set of rule and cost parameters to apply across the experiment that would allow for some differentiation of results, while containing the number of treatments required. As described in Chapter 3 of this report, the researchers chose to fix the supplier options across all treatments as one way to minimise the number of treatments considered in the experiment.

Appendix B.2 below details the functions used to estimate the expected returns for various supplier strategies under different cost and rule structures. The number of runs (the "trials" variable in the MATLAB code) was set to 500.[102] The code in this appendix references both the CSP-1 an CSP-3 rule. A feature of this program is that it assumes a return of 1 for selling the goods in the domestic market, as done in Rossiter and Hester (2017) as a normalisation. Multiplying these matrices by a factor of 20 gives the types of results in Figures 8 and 9 in Chapter 3.5 of this report.

The code in Appendix B.3 was a sensitivity test for how the predicted "optimal" strategy in this framework might change when moving from a risk-neutral to a risk-averse subject. The results are not reported in this document, mainly because they did not appear to change much for our choice of model parameters in the experiment.[103] This provided a further robustness check on predicted behaviour results summarised in Table 7 of the *Final Report*.

B.2. Implementing simulations for risk-neutral importers

B.2.1. MATLAB function for simulating the CSP-1 algorithm predictions

```
function [CSP1mat, constmat] = CSP1_suppsolve(tablook, CN, MFlow, ...
   MFtrue, MFhigh, omega, kappa, chi, gamma, pdetect, consign, trials)

% Note: this code requires the following parameters to be specified outside
% the model:
% CN - clearance number required to revert to monitoring mode
% TC - the number of items required to be passed consecutively in tight
% census mode
% HT - the number of items required to be passed consecutively after a
% failure for the counter to be reset so another failure doesn't revert to
% census mode again
% MFlow - the lower bound on the monitoring fraction stated to importers
% MFtrue - the true monitoring fraction in monitoring mode
% MFhigh - the upper bound on the monitoring fraction stated to importers
```

[102] Varying this number up or slightly down did not seem to radically affect the simulation results. In particular, the "optimal" strategy did not seem to change by doing this type of adjustment.

[103] In only a few treatments, increasing the risk aversion parameter in the constant relative risk aversion (CRRA) function resulted in the simulation results changing to favour lower-risk suppliers (i.e. supplier D) more noticeably.

```
% omega - cost incurred by importer from being inspected
% kappa - cost of treatment for importer from failing inspection
% chi - cost incurred by inspector from completing an inspection
% gamma - cost of a consignment with biosecurity risk material
% pdetect - probability that biosecurity hazards are detected in a shipment
% that contains them
% consign - number of consignments used in simulation
% trials - number of iterations used to construct expected payoff function

% NOTE: implementing this program transforms the payoff structure to ensure
% all payoffs are at least one to avoid undefined results

% Provides a way to evaluate the "fitness" associated with a CSP-1
% algorithm with given inputs associated with tablook
% The columns of tablook indicate a different potential supplier
% The rows of tablook represent:
% Row 1 - cost per unit of importing from there
% Row 2 - the probability of having biosecurity risk material in a given
% consignment

% Design matrix to capture number of options available and payoffs from
% different options
% First column of matrix lists supplier choice in census
% Second column of matrix lists supplier choice in monitoring
% Third column of matrix lists supplier cost in census
% Fourth column of matrix lists supplier cost in monitoring
% Fifth column of matrix lists the approach rate in census
% Sixth column of matrix lists the approach rate in monitoring
% Seventh column of matrix lists average importer payoff under MFlow
% Eighth column of matrix lists average inspector payoff under MFlow
% Ninth column of matrix lists average importer payoff under MFtrue
% Tenth column of matrix lists average inspector payoff under MFtrue
% Eleventh column of matrix lists average importer payoff under MFhigh
% Twelfth column of matrix lists average inspector payoff under MFhigh
nosupp = size(tablook,2);
CSP1mat = [kron((1:nosupp)',ones(nosupp,1)) kron(ones(nosupp,1),(1:nosupp)') ...
   kron(tablook(1,:)',ones(nosupp,1)) kron(ones(nosupp,1),tablook(1,:)') ...
   kron(tablook(2,:)',ones(nosupp,1)) kron(ones(nosupp,1),tablook(2,:)') ...
   zeros(nosupp^2,6)];
MFmat = [MFlow MFtrue MFhigh];

% For constant probability of inspection (either 1 or MF), matrix given by
% First column of matrix lists supplier choice
% Second column of matrix lists supplier cost
% Third column of matrix lists approach rate
% Fourth column of matrix lists average importer payoff in mandatory
% inspection
% Fifth column of matrix lists average inspector payoff in mandatory
% inspection
% Sixth column of matrix lists average importer payoff in inspecting with
```

```
% probability MF
% Seventh column of matrix lists average inspector payoff in inspecting
% with probability MF
constmat = [(1:nosupp)' tablook' zeros(nosupp, 4)];

% Do iterations over options for CSP1 scheme
for j = 1 : nosupp^2
   % Extract probability of failure from matrix
   probfail = CSP1mat(j, 5:6);
   % Extract variable supplier cost level
   abatetec = CSP1mat(j, 3:4);

   for m = 1 : 3
      % Construct vector to retain values for expected payoff - first
      % column for importer, second column for inspector
      simvec = zeros(trials, 2);
      MFval = MFmat(1, m);

      for k = 1 : trials
         chainseq = 1;
         algmode = 1;
         for c = 1 : consign
            % Whether biosecurity hazards present & whether detected
            biosec = (rand < probfail(1, algmode));
            detected = biosec * (rand < pdetect);
            % Enhanced inspection mode or tight census mode
            if chainseq < CN + 1
               % Inspector payoff from inspection - based on actual incursion
               simvec(k, 2) = simvec(k, 2) + ...
                  (chi + (biosec - detected) * (gamma - chi));
               % Importer payoff - only incurs treatment cost if detected
               simvec(k, 1) = simvec(k, 1) + ...
                  (abatetec(1, 1) + omega + kappa * detected - 1);
               % Return to state 1 if biosecurity hazards detected or keep
               % proceeding
               chainseq = 1 + (1 - detected) * chainseq;
               algmode = 1 + (chainseq == CN + 1);
               % Monitoring mode
            else
               % Not inspected
               if rand > MFval
                  % Importer payoff
                  simvec(k, 1) = simvec(k, 1) + abatetec(1, 2) - 1;
                  % Inspector payoff
                  simvec(k, 2) = simvec(k, 2) + biosec * gamma;
               else
                  % Subject to inspection
                  simvec(k, 2) = simvec(k, 2) + ...
                     (chi + (biosec - detected) * (gamma - chi));
                  % Importer payoff
```

```
                simvec(k, 1) = simvec(k,1) + ...
                   (abatetec(1, 2) + omega + kappa * detected - 1);
                % Determines where next step goes
                chainseq = 1 + (1 - detected) * CN;
                algmode = 1 + (chainseq == CN + 1);
             end
          end
       end
    end
    % Construct mean payoff matrix
    CSP1mat(j,5+2*m:6+2*m) = -mean(simvec, 1);
  end
end

% Mandatory inspection case
for j = 1 : nosupp
  % Extract probability of failure from matrix
  probfail = tablook(2, j);
  % Extract variable supplier cost level
  abatetec = tablook(1, j);
  % Construct vector to retain values for expected payoff - first
  % column for importer, second column for inspector
  simvec = zeros(trials, 2);
  for k = 1 : trials
    for c = 1 : consign
      % Whether biosecurity hazards present & whether detected
      biosec = (rand < probfail);
      detected = biosec * (rand < pdetect);
      % Inspector payoff from inspection - based on actual incursion
      simvec(k, 2) = simvec(k, 2) + ...
         (chi + (biosec - detected) * (gamma - chi));
      % Importer payoff - only incurs treatment cost if detected
      simvec(k, 1) = simvec(k, 1) + ...
         (abatetec + omega + kappa * detected - 1);
    end
  end
  % Construct mean payoff matrix
  constmat(j,4:5) = -mean(simvec, 1);
end

% Random inspection with fixed frequency case
for j = 1 : nosupp
  % Extract probability of failure from matrix
  probfail = tablook(2, j);
  % Extract variable supplier cost level
  abatetec = tablook(1, j);
  % Construct vector to retain values for expected payoff - first
  % column for importer, second column for inspector
  simvec = zeros(trials, 2);
  for k = 1 : trials
```

```
      for c = 1 : consign
         % Whether biosecurity hazards present & whether detected
         biosec = (rand < probfail);
         detected = biosec * (rand < pdetect);
         % Not inspected case first
         if rand > MFtrue
            % Importer payoff
            simvec(k, 1) = simvec(k, 1) + abatetec - 1;
            % Inspector payoff
            simvec(k, 2) = simvec(k, 2) + biosec * gamma;
         else
            % Subject to inspection
            simvec(k, 2) = simvec(k, 2) + ...
               (chi + (biosec - detected) * (gamma - chi));
            % Importer payoff
            simvec(k, 1) = simvec(k,1) + ...
               (abatetec + omega + kappa * detected - 1);
         end
      end
   end
   % Construct mean payoff matrix
   constmat(j,6:7) = -mean(simvec, 1);
end
```

B.2.2. MATLAB function for simulating the CSP-3 algorithm prediction

```
function [CSP1mat, constmat] = CSP3_suppsolve(tablook, CN, TC, HT, ...
   MFlow, MFtrue, MFhigh, omega, kappa, chi, gamma, pdetect, consign, trials)

% Note: this code requires the following parameters to be specified outside
% the model:
% CN - clearance number required to revert to monitoring mode
% TC - the number of items required to be passed consecutively in tight
% census mode
% HT - the number of items required to be passed consecutively after a
% failure for the counter to be reset so another failure doesn't revert to
% census mode again
% MFlow - the lower bound on the monitoring fraction stated to importers
% MFtrue - the true monitoring fraction in monitoring mode
% MFhigh - the upper bound on the monitoring fraction stated to importers
% omega - cost incurred by importer from being inspected
% kappa - cost of treatment for importer from failing inspection
% chi - cost incurred by inspector from completing an inspection
% gamma - cost of a consignment with biosecurity risk material
% pdetect - probability that biosecurity hazards are detected in a shipment
% that contains them
% consign - number of consignments used in simulation
% trials - number of iterations used to construct expected payoff function

% NOTE: implementing this program transforms the payoff structure to ensure
% all payoffs are at least one to avoid undefined results
```

```
% Provides a way to evaluate the "fitness" associated with a CSP-1
% algorithm with given inputs associated with tablook
% The columns of tablook indicate a different potential supplier
% The rows of tablook represent:
% Row 1 - cost per unit of importing from there
% Row 2 - the probability of having biosecurity risk material in a given
% consignment

% Design matrix to capture number of options available and payoffs from
% different options
% First column of matrix lists supplier choice in census
% Second column of matrix lists supplier choice in monitoring
% Third column of matrix lists supplier cost in census
% Fourth column of matrix lists supplier cost in monitoring
% Fifth column of matrix lists the approach rate in census
% Sixth column of matrix lists the approach rate in monitoring
% Seventh column of matrix lists average importer payoff under MFlow
% Eighth column of matrix lists average inspector payoff under MFlow
% Ninth column of matrix lists average importer payoff under MFtrue
% Tenth column of matrix lists average inspector payoff under MFtrue
% Eleventh column of matrix lists average importer payoff under MFhigh
% Twelfth column of matrix lists average inspector payoff under MFhigh
nosupp = size(tablook,2);
CSP1mat = [kron((1:nosupp)',ones(nosupp,1)) kron(ones(nosupp,1),(1:nosupp)') ...
   kron(tablook(1,:)',ones(nosupp,1)) kron(ones(nosupp,1),tablook(1,:)') ...
   kron(tablook(2,:)',ones(nosupp,1)) kron(ones(nosupp,1),tablook(2,:)') ...
   zeros(nosupp^2,6)];
MFmat = [MFlow MFtrue MFhigh];

% For constant probability of inspection (either 1 or MF), matrix given by
% First column of matrix lists supplier choice
% Second column of matrix lists supplier cost
% Third column of matrix lists approach rate
% Fourth column of matrix lists average importer payoff in mandatory
% inspection
% Fifth column of matrix lists average inspector payoff in mandatory
% inspection
% Sixth column of matrix lists average importer payoff in inspecting with
% probability MF
% Seventh column of matrix lists average inspector payoff in inspecting
% with probability MF
constmat = [(1:nosupp)' tablook' zeros(nosupp, 4)];

% Do iterations over options for CSP1 scheme
for j = 1 : nosupp^2
   % Extract probability of failure from matrix
   probfail = CSP1mat(j, 5:6);
   % Extract variable supplier cost level
   abatetec = CSP1mat(j, 3:4);
```

```
for m = 1 : 3
   % Construct vector to retain values for expected payoff - first
   % column for importer, second column for inspector
   simvec = zeros(trials, 2);
   MFval = MFmat(1, m);

   for k = 1 : trials
      chainseq = 1;
      algmode = 1;
      for c = 1 : consign
         % Whether biosecurity hazards present & whether detected
         biosec = (rand < probfail(1, 2- mod(algmode,2)));
         detected = biosec * (rand < pdetect);
         % Enhanced inspection mode or tight census mode
         if chainseq < CN + 1 || (chainseq > CN + 1 && ...
            chainseq <= CN + TC + 1)
            % Inspector payoff from inspection - based on actual incursion
            simvec(k, 2) = simvec(k, 2) + ...
               (chi + (biosec - detected) * (gamma - chi));
            % Importer payoff - only incurs treatment cost if detected
            simvec(k, 1) = simvec(k, 1) + ...
               (abatetec(1, 1) + omega + kappa * detected - 1);
            % Return to state 1 if biosecurity hazards detected or keep
            % proceeding
            chainseq = 1 + (1 - detected) * chainseq;
            algmode = 1 + (chainseq == CN + 1) + ...
               (chainseq > CN + 1) + (chainseq > CN + TC + 1);
            % Monitoring mode
         else
            % Not inspected
            if rand > MFval
               % Importer payoff
               simvec(k, 1) = simvec(k, 1) + abatetec(1, 2) - 1;
               % Inspector payoff
               simvec(k, 2) = simvec(k, 2) + biosec * gamma;
            else
               % Subject to inspection
               simvec(k, 2) = simvec(k, 2) + ...
                  (chi + (biosec - detected) * (gamma - chi));
               % Importer payoff
               simvec(k, 1) = simvec(k,1) + ...
                  (abatetec(1, 2) + omega + kappa * detected - 1);
               % Determines where next step goes
               if chainseq == CN + 1
                  chainseq = chainseq + detected;
                  algmode = 2 + detected;
               else
                  chainseq = detected + (1 - detected) * ...
                     (chainseq + 1 - (chainseq == CN+HT+1) * (HT+1));
```

```
                    algmode = 1 + (chainseq == CN + HT + 1) + ...
                      3 * (chainseq < CN + HT + 1) * (chainseq > CN + 1);
                end
              end
            end
          end
        end
        % Construct mean payoff matrix
        CSP1mat(j,5+2*m:6+2*m) = -mean(simvec, 1);
    end
end

% Mandatory inspection case
for j = 1 : nosupp
    % Extract probability of failure from matrix
    probfail = tablook(2, j);
    % Extract variable supplier cost level
    abatetec = tablook(1, j);
    % Construct vector to retain values for expected payoff - first
    % column for importer, second column for inspector
    simvec = zeros(trials, 2);
    for k = 1 : trials
        for c = 1 : consign
            % Whether biosecurity hazards present & whether detected
            biosec = (rand < probfail);
            detected = biosec * (rand < pdetect);
            % Inspector payoff from inspection - based on actual incursion
            simvec(k, 2) = simvec(k, 2) + ...
                (chi + (biosec - detected) * (gamma - chi));
            % Importer payoff - only incurs treatment cost if detected
            simvec(k, 1) = simvec(k, 1) + ...
                (abatetec + omega + kappa * detected - 1);
        end
    end
    % Construct mean payoff matrix
    constmat(j,4:5) = -mean(simvec, 1);
end

% Random inspection with fixed frequency case
for j = 1 : nosupp
    % Extract probability of failure from matrix
    probfail = tablook(2, j);
    % Extract variable supplier cost level
    abatetec = tablook(1, j);
    % Construct vector to retain values for expected payoff - first
    % column for importer, second column for inspector
    simvec = zeros(trials, 2);
    for k = 1 : trials
        for c = 1 : consign
            % Whether biosecurity hazards present & whether detected
```

```
        biosec = (rand < probfail);
        detected = biosec * (rand < pdetect);
        % Not inspected case first
        if rand > MFtrue
           % Importer payoff
           simvec(k, 1) = simvec(k, 1) + abatetec - 1;
           % Inspector payoff
           simvec(k, 2) = simvec(k, 2) + biosec * gamma;
        else
           % Subject to inspection
           simvec(k, 2) = simvec(k, 2) + ...
              (chi + (biosec - detected) * (gamma - chi));
           % Importer payoff
           simvec(k, 1) = simvec(k,1) + ...
              (abatetec + omega + kappa * detected - 1);
        end
     end
  end
  % Construct mean payoff matrix
  constmat(j,6:7) = -mean(simvec, 1);
end
```

B.3. Implementing simulations for importers with constant relative risk aversion

B.3.1. Constant relative risk aversion (CRRA) utility function payoff conversion

```
function [imptab, regtab] = CRRAutil(tablook, omega, kappa, ...
  chi, gamma, impra, regra)

% Transforms payoff structure into form of CRRA utility function to assess
% effect of risk aversion on the solution

% The columns of tablook indicate a different potential supplier
% The rows of tablook represent:
% Row 1 - cost per unit of importing from there
% Row 2 - the probability of having biosecurity risk material in a given
% consignment
% Only row 1 is used in this program

% omega - cost incurred by importer from being inspected
% kappa - cost of treatment for importer from failing inspection
% chi - cost incurred by inspector from completing an inspection
% gamma - cost of a consignment with biosecurity risk material
% impra - risk aversion parameter for CRRA utility for the importer
% regra - risk aversion parameter for CRRA utility for the inspection

% Raw payoff matrices for the importer
% Row 1 - payoff when not inspected for each potential supplier
% Row 2 - payoff when inspected and determined to not have biosecurity risk
% material for each potential supplier
```

```
% Row 3 - payoff when inspected and found to have biosecurity risk material
% and therefore requires treatment for each potential supplier
imptab = [1 - tablook(1,:); 1 - (omega + tablook(1,:)); ...
   1 - (omega + kappa + tablook(1,:))];
% Translate payoffs to ensure minimum payoff is at least 0.1 - to allow for
% extension
imptab = imptab - min(min(imptab)) + 0.1;

% Convert using CRRA utility
if impra == 1
   imptab = log(imptab);
else
   imptab = imptab.^(1 - impra) / (1 - impra);
end

% Inspection payoff vector
% Row 1 - payoff when not inspected & consignment clean
% Row 2 - consignment inspected
% Row 3 - consignment not inspected but contains biosecurity risk material
% Row 4 - consignment inspected, taken as passed inspection but actually
% contains biosecurity risk material
regtab = [0; -chi; -gamma; -(chi + gamma)];

% Translate payoffs to ensure minimum payoff is at least 0.1 - to allow for
% extension
regtab = regtab - min(regtab) + 0.1;

% Convert using CRRA utility
if regra == 1
   regtab = log(regtab);
else
   regtab = regtab.^(1 - regra) / (1 - regra);
end
```

B.3.2. MATLAB function for simulating the CSP-1 algorithm prediction under CRRA utility

```
function [CSP1mat, constmat] = CSP1_suppsolve_crra(tablook, CN, MFlow, ...
   MFtrue, MFhigh, omega, kappa, chi, gamma, pdetect, consign, trials, ...
   impra, regra, scalefac)

% Note: this code requires the following parameters to be specified outside
% the model:
% CN - clearance number required to revert to monitoring mode
% TC - the number of items required to be passed consecutively in tight
% census mode
% HT - the number of items required to be passed consecutively after a
% failure for the counter to be reset so another failure doesn't revert to
% census mode again
% MFlow - the lower bound on the monitoring fraction stated to importers
% MFtrue - the true monitoring fraction in monitoring mode
```

```
% MFhigh - the upper bound on the monitoring fraction stated to importers
% omega - cost incurred by importer from being inspected
% kappa - cost of treatment for importer from failing inspection
% chi - cost incurred by inspector from completing an inspection
% gamma - cost of a consignment with biosecurity risk material
% pdetect - probability that biosecurity hazards are detected in a shipment
% that contains them
% consign - number of consignments used in simulation
% trials - number of iterations used to construct expected payoff function
% impra - risk aversion parameter for CRRA utility for the importer
% regra - risk aversion parameter for CRRA utility for the inspection

% NOTE: implementing this program transforms the payoff structure to ensure
% all payoffs are at least one to avoid undefined results

% Provides a way to evaluate the "fitness" associated with a CSP-1
% algorithm with given inputs associated with tablook
% The columns of tablook indicate a different potential supplier
% The rows of tablook represent:
% Row 1 - cost per unit of importing from there
% Row 2 - the probability of having biosecurity risk material in a given
% consignment

% Design matrix to capture number of options available and payoffs from
% different options
% First column of matrix lists supplier choice in census
% Second column of matrix lists supplier choice in monitoring
% Third column of matrix lists the approach rate in census
% Fourth column of matrix lists the approach rate in monitoring
% Fifth column of matrix lists average importer payoff under MFlow
% Sixth column of matrix lists average inspector payoff under MFlow
% Seventh column of matrix lists average importer payoff under MFtrue
% Eighth column of matrix lists average inspector payoff under MFtrue
% Ninth column of matrix lists average importer payoff under MFhigh
% Tenth column of matrix lists average inspector payoff under MFhigh
nosupp = size(tablook,2);
CSP1mat = [kron((1:nosupp)',ones(nosupp,1)) ...
   kron(ones(nosupp,1),(1:nosupp)') ...
   kron(tablook(2,:)',ones(nosupp,1)) ...
   kron(ones(nosupp,1),tablook(2,:)') ...
   zeros(nosupp^2,6)];
MFmat = [MFlow MFtrue MFhigh];

% For constant probability of inspection (either 1 or MF), matrix given by
% First column of matrix lists supplier choice
% Second column of matrix lists approach rate
% Third column of matrix lists average importer payoff in mandatory
% inspection
% Fourth column of matrix lists average inspector payoff in mandatory
% inspection
```

```
% Fifth column of matrix lists average importer payoff in inspecting with
% probability MF
% Sixth column of matrix lists average inspector payoff in inspecting
% with probability MF
constmat = [(1:nosupp)' tablook(2,:)' zeros(nosupp, 4)];

% Compute payoff matrices for importer and inspector (regulator) taking
% into account risk preferences from CRRA utility
% Structure of imptab: adjusted payoff matrices for the importer
% Row 1 - payoff when not inspected for each potential supplier
% Row 2 - payoff when inspected and determined to not have biosecurity risk
% material for each potential supplier
% Row 3 - payoff when inspected and found to have biosecurity risk material
% and therefore requires treatment for each potential supplier
% Structure of regtab: inspection payoff vector
% Row 1 - payoff when not inspected & consignment clean
% Row 2 - consignment inspected
% Row 3 - consignment not inspected but contains biosecurity risk material
% Row 4 - consignment inspected, taken as passed inspection but actually
% contains biosecurity risk material
[imptab, regtab] = CRRAutil(tablook, omega, kappa, ...
   chi, gamma, impra, regra);

% Do iterations over options for CSP1 scheme
for j = 1 : nosupp^2
   % Extract probability of failure from matrix
   probfail = CSP1mat(j, 3:4);
   % Extract payoffs for various alternative supplier choices
   payimp = [imptab(:,CSP1mat(j,1)) imptab(:,CSP1mat(j,2))];

   for m = 1 : 3
      % Construct vector to retain values for expected payoff - first
      % column for importer, second column for inspector
      simvec = zeros(trials, 2);
      MFval = MFmat(1, m);

      for k = 1 : trials
         chainseq = 1;
         algmode = 1;
         for c = 1 : consign
            % Whether biosecurity hazards present & whether detected
            biosec = (rand < probfail(1, algmode));
            detected = biosec * (rand < pdetect);
            notdet = biosec - detected;
            % Enhanced inspection mode or tight census mode
            if chainseq < CN + 1
               % Inspector payoff from inspection - based on actual incursion
               simvec(k, 2) = simvec(k, 2) + ...
                  (notdet == 0) * regtab(2,1) + ...
                  (notdet == 1) * regtab(4,1);
```

```
            % Importer payoff - only incurs treatment cost if detected
            simvec(k, 1) = simvec(k, 1) + ...
               (detected == 0) * payimp(2,1) + ...
               (detected == 1) * payimp(3,1);
            % Return to state 1 if biosecurity hazards detected or keep
            % proceeding
            chainseq = 1 + (1 - detected) * chainseq;
            algmode = 1 + (chainseq == CN + 1);
            % Monitoring mode
         else
            % Not inspected
            if rand > MFval
               % Importer payoff
               simvec(k, 1) = simvec(k, 1) + payimp(1,2);
               % Inspector payoff
               simvec(k, 2) = simvec(k, 2) + ...
                  (biosec == 0) * regtab(1,1) + ...
                  (biosec == 1) * regtab(3,1);
            else
               % Subject to inspection - inspector payoff
               simvec(k, 2) = simvec(k, 2) + ...
                  (notdet == 0) * regtab(2,1) + ...
                  (notdet == 1) * regtab(4,1);
               % Importer payoff
               simvec(k, 1) = simvec(k,1) + ...
                  (detected == 0) * payimp(2,2) + ...
                  (detected == 1) * payimp(3,2);
               % Determines where next step goes
               chainseq = 1 + (1 - detected) * CN;
               algmode = 1 + (chainseq == CN + 1);
            end
         end
      end
      % Construct mean payoff matrix
      CSP1mat(j,3+2*m:4+2*m) = mean(simvec, 1) * scalefac;
   end
end

% Mandatory inspection case
for j = 1 : nosupp
   % Extract probability of failure from matrix
   probfail = tablook(2, j);
   % Extract variable supplier cost level
   payimp = imptab(:, j);
   % Construct vector to retain values for expected payoff - first
   % column for importer, second column for inspector
   simvec = zeros(trials, 2);
   for k = 1 : trials
      for c = 1 : consign
```

```
        % Whether biosecurity hazards present & whether detected
        biosec = (rand < probfail);
        detected = biosec * (rand < pdetect);
        notdet = biosec - detected;
        % Inspector payoff from inspection - based on actual incursion
        simvec(k, 2) = simvec(k, 2) + (notdet == 0) * regtab(2,1) + ...
           (notdet == 1) * regtab(4,1);
        % Importer payoff - only incurs treatment cost if detected
        simvec(k, 1) = simvec(k, 1) + ...
           (detected == 0) * payimp(2,1) + ...
           (detected == 1) * payimp(3,1);
     end
  end
  % Construct mean payoff matrix
  constmat(j,3:4) = mean(simvec, 1) * scalefac;
end

% Random inspection with fixed frequency case
for j = 1 : nosupp
  % Extract probability of failure from matrix
  probfail = tablook(2, j);
  % Extract variable supplier cost level
  payimp = imptab(:, j);
  % Construct vector to retain values for expected payoff - first
  % column for importer, second column for inspector
  simvec = zeros(trials, 2);
  for k = 1 : trials
     for c = 1 : consign
        % Whether biosecurity hazards present & whether detected
        biosec = (rand < probfail);
        detected = biosec * (rand < pdetect);
        notdet = biosec - detected;
        % Not inspected case first
        if rand > MFtrue
           % Importer payoff
           simvec(k, 1) = simvec(k, 1) + payimp(1,1);
           % Inspector payoff
           simvec(k, 2) = simvec(k, 2) + ...
              (biosec == 0) * regtab(1,1) + ...
              (biosec == 1) * regtab(3,1);
        else
           % Subject to inspection - inspector payoff
           simvec(k, 2) = simvec(k, 2) + ...
              (notdet == 0) * regtab(2,1) + ...
              (notdet == 1) * regtab(4,1);
           % Importer payoff
           simvec(k, 1) = simvec(k,1) + ...
              (detected == 0) * payimp(2,1) + ...
              (detected == 1) * payimp(3,1);
        end
```

```
        end
    end
    % Construct mean payoff matrix
    constmat(j,5:6) = mean(simvec, 1) * scalefac;
end
```