Modelling the Human and Technological Costs and Benefits of USB Memory Stick Security

Adam Beautement[1], Robert Coles[2], Jonathan Griffin[3], Christos Ioannidis[4], Brian Monahan[3], David Pym[34*], Angela Sasse[1] and Mike Wonham[3]

[1]UCL, [2]Merrill Lynch, [3]HP Labs, [4]University of Bath

Abstract Organizations deploy systems technologies in order to support their operations and achieve their business objectives. In so doing, they encounter tensions between the confidentiality, integrity, and availability of information, and must make investments in information security measures to address these concerns. We discuss how a macroeconomics-inspired model, analogous to models of interest rate policy used by central banks, can be used to understand trade-offs between investments against threats to confidentiality and availability. We investigate how such a model might be formulated by constructing a process model, based on empirically obtained data, of the use of USB memory sticks by employees of a financial services company.

1 Introduction

Organizations deploy systems technologies in order to support their operations and achieve their business objectives. In so doing, they encounter tensions between the confidentiality, integrity, and availability of information.

In formulating security policies that are intended to resolve such tensions to the organizations' satisfaction, people (e.g., CEOs, CIOs, CISOs, security managers) with responsibility for information and systems security face the following two problems:

1. Poor economic understanding of how to formulate, resource, measure, and value security policies; and
2. Poor organizational understanding of the attitudes of users to both information and systems security and of their responses to imposed security policies (see, for example, the UK Foresight 'Cyber Trust and Crime Prevention' report (Office of Science and Technology 2004)).

* Corresponding author.

M.E. Johnson (ed.), *Managing Information Risk and the Economics of Security*,
DOI: 10.1007/978-0-387-09762-6_7,

Consequently, the effectiveness and value of the policies with which users are expected to comply are very difficult to assess, as are the corresponding investment decisions (Anderson 2001; Anderson and Moore 2006). We believe that, in order to assess the effectiveness and value of security investments in a system, be they in people, process, or technology, it is necessary to have a conceptualization (i.e., a model) of the system, including its users, and its economic environment.

In this work, we present an entirely novel approach to the problem of modelling the economic effectiveness of implementing security policies within an organization. The following are the key components of our approach:

- We test the hypothesis that there is a trade-off between the components of investments in information security that address confidentiality and availability (for our present purposes, we suppress integrity);
- For now, we capture primarily conceptually rather than mathematically, the trade-off between availability and confidentiality using a model inspired by a macroeconomic model of the Central Bank Problem (Ruge-Murcia 2001; 2003). Our approach, which considers aggregate values of confidentiality and availability under variation in investment, stands in contrast to the microeconomic approaches described by Gordon and Loeb (Gordon and Loeb 2002; 2006);
- Rather than provide a detailed mathematical formulation, which at this stage in our investigation we are not ready to formulate, we conduct an empirical study together with a (rigorously structured) simulation based on the empirical data and the processes executed by the system. Our simulations embody the dynamics of the conceptual model;
- Our empirical data is obtained from semi-structured interviews with staff at two organizations, a financial services company and a research organization, with a focus here on the financial services organization;
- We demonstrate the use of the model to explore the utility of trade-offs between availability and confidentiality.

The results of our study, and variations upon it, will inform our efforts to design and calibrate economic models of the kind we discuss.

The remainder of the chapter is structured as follows: In Section 2, we explain the form of the economic model of the response of confidentiality and availability to security investments that is of interest to us; in Section 3, we explain how we have obtained our initial empirical data; in Section 4, we explain the key features of our process model of the use of USB memory sticks and, in Section 5, we explain how this model is realized in our systems modelling language, Demos2k (Demos2k); in Section 6, we explain our experimental study, including its relationship to the economic model we sketch in Section 2; and finally, in Section 7, we explain how we intend to pursue this work, explaining the directions of empirical study, process modelling, and economic modelling. We also include two appendices, one containing a summary of the empirical data and one containing the code for our (executable) model; both are available at http://weis2008.econinfosec.org.

2 The Central Bank Problem and Information Security

A well-known problem in macroeconomics concerns the setting of interest rates by a central bank in order to manage, say, inflation and (un)employment. The basic model derives from a line of work including Barro and Gordon (1983), Taylor (1993), and Nobay and Peel (2003).

In very brief summary, for readers who may be unfamiliar with the background, the basic setup of the model is as follows (Ruge-Murcia 2001; 2003):

- Inflation and unemployment are related as

$$u_t = u_t^n - \lambda\left(\pi_t - \pi_t^e\right) + \eta_t,$$

for $\lambda>0$, where u_t, u_t^n and π_t are, respectively, the rates of unemployment, natural (or target) unemployment, and inflation; π_t^e is the (public) forecast of inflation at time t, constructed at time t–1, determined rationally as

$$\pi_t^e = E_{t-1}\pi_t,$$

where E_{t-1} is the expectation conditional on the set of all relevant information available at time t–1, denoted I_{t-1}; η_t is an aggregate supply disturbance;

- The natural (or target) rate of unemployment evolves over time, with Δu_t^n depending on the $\Delta u_{t-k}^n s$;
- The central bank affects the rate of inflation via a policy instrument, such as a base interest rate. Such an instrument is imperfect, with imperfections represented by the error term ε_t in the following equation, in which $i_t \in I_{t-1}$:

$$\pi_t = i_t + \varepsilon_t;$$

- The central bank's preferences for inflation and unemployment are captured by a utility, or loss, function of the following form:

$$U\left(\pi_t, u_t\right) = \frac{\phi}{2}\left(\pi_t - \pi_t^*\right)^2 + \frac{\phi}{\gamma^2}\left(\exp\left(\gamma\left(u_t - u_t^*\right)\right) - \gamma\left(u_t - u_t^*\right) - 1\right),$$

where π_t^* and u_t^*, respectively, are the target rates of inflation and unemployment, and ϕ is a parameter; γ is a non-zero real. Here the target unemployment rate is the expected (natural) rate of unemployment:

$$u_t^* = E_{t-1}\left(u_t^n\right).$$

It is assumed that the target inflation, π_t^*, can be approximated by a constant term (Ruge-Murcia 2001; 2003).

Note that the utility function taken in this setup employs the linex function (Varian 1974; Zellner 1986; Clatworthy et al. 2006), of the form

$$g\left(x\right) = \left(\exp\left(\alpha x\right) - \alpha x - 1\right) / \alpha^2$$

where α is a parameter. In comparison with the use of a quadratic utility function, the linex function admits asymmetry whilst retaining the quadratic as the special (limit) case when α tends to zero.

We argue that a form of the central bank problem (model) can be deployed to explain trade-offs in investments in information security. In our present case, we are concerned with the trade-off between availability and confidentiality, in the particular setting of the overall availability of information derived from the use of USB memory sticks set against the overall increased exposure of confidential information that is a consequence of their use. The analogy goes as follows:

- Availability and confidentiality, respectively, correspond to inflation and unemployment. The policy instrument is the level of investment in information security countermeasures;
- Availability and confidentiality are related as follows:
  - As availability increases, the potential for exposures increases, and confidentiality decreases. Confidentiality is also reduced by increased levels of threat to confidentiality
    $$C = -\lambda A + \varepsilon_C ,$$
    where λ is a parameter and ε_C is a non-decreasing stochastic process (so expectation is non-zero) for the threat to confidentiality;
  - Availability depends both on the level of investment in information security, negatively in the case of the study discussed in this chapter, and on the level of threat to availability
    $$A = -\Psi I + \varepsilon_A ,$$
    where the instrument I is security investment or, perhaps, system complexity, Ψ is a (possibly negative) parameter and ε_A is a non-decreasing stochastic process for the threat to availability. More generally, we might also require a term in changes ΔI in the instrument I, with various dependencies;
- For utility, in terms of expectations, we might take, for example,
  $$E\left(U\left(C,A\right)\right) = E\left(\left(\exp\left[\alpha A\right] - \alpha A - 1\right)/\alpha^2 + \frac{\phi}{2}C^2\right),$$
  where ϕ is a parameter, as before;
- Such a formulation does have analytic solutions for I, in terms of expectation, of the form
  $$I = E\left(\frac{1}{\Psi}\left[\varepsilon_A - \frac{\varepsilon C}{\lambda} - \frac{1}{\alpha\lambda^2\phi} + \text{ProductLog}\left[\frac{\exp\left(\frac{\alpha\varepsilon_C}{\lambda} + \frac{1}{\lambda^2\phi}\right)}{\lambda^2\phi}\right]\right]\right),$$
  where, as in Mathematica (2008), ProductLog[z] is a solution for w in $z = w\exp(w)$. A discussion of this solution and its significance is beyond our present scope, as is a discussion of a multi-period model.

As we have remarked, in the context of information systems, the instrument I might be a measure of investment in information security, or a measure of the complexity of the system. For an example of the latter, we might take a 'complexity parameter', $x \in [0,1)$, and then take $I=1/(1-x)$. Then if $x=0$, we have a maximally simple system (a single unit) and, as x approaches 1, and so I approaches infinity, we can obtain an arbitrarily complex system.

In business contexts, systems users who have access to confidential and business-critical information make widespread use of USB memory sticks. They do so for good reasons: these devices efficiently enable data transfer between all manner of business colleagues and partners. The use of these devices also exposes organizations to risks of losses of confidential data, owing to their capability to transfer all kinds of data conveniently and cheaply to anyone capable of receiving it. Thus there is a trade-off between availability and confidentiality (we suppress consideration of integrity issues in this context, where it can be argued that they are minor), and there is an incentive incompatibility between the users of the systems and owners of the policies.

In this chapter, we study the use of USB memory sticks by the staff of a financial services firm, in the context of a model of the form discussed above. We do not attempt to reify such a model analytically, even at this level of detail. Rather, we demonstrate the dynamics of a simple instance using an executable model of the system of USB users using a process model.

The model, built on the basis of empirically obtained data, executes processes that track availability and breaches of confidentiality under specified levels of security investment. In assessing our experimental results within the executable model, we employ, for illustrative purposes, perhaps the simplest form of utility function that might possibly be useful:

$$U(C, A) = \alpha(A - \beta C),$$

where α and β are parameters; the details of the choices here are explained in Section 6.

3 An Empirical Study

To obtain an empirical basis for our model, we conducted a study to elicit factors that contribute to corporate and individual security cost. One of the academic researchers conducted 17 in-depth interviews with security staff, employees, and managers in the two companies that are partners in this research project. The interviews remained anonymous.

The interviews were semi-structured, exploring

- the tasks and responsibilities of interviewees,
- their perception of the risks facing the company,

- their attitudes to the company's security polices and security measures, and
- the perceived impact of security measures on individuals' tasks and responsibilities, as well as company productivity.

Whilst the interviews covered a range of security policies and measures, all interviewees were asked about one specific security problem: USB sticks. They were asked

- if they used USB sticks (all did),
- how they used them as part of their tasks and responsibilities,
- about the relationship between the risks facing their company, and their USB stick usage,
- if whether any of their USB stick usage contravened the company's security policies, and if so,
- why they thought contravening the security policy was justified.

We suggested the company was considering making the use of encrypted USB sticks mandatory (for the financial services company, this was actually the case), and asked interviewees to

- explore the cost and benefits of such a policy for the company, and
- explain the cost and benefit for them and their tasks and responsibilities.

The interviews were transcribed and analyzed using techniques from Grounded Theory. Grounded Theory (Strauss and Corbine 1990) is a qualitative data analysis method widely used in social sciences, which allows identification of salient concepts and relationships between them. Over the past 10 years, the method has been successfully applied to model user perceptions and attitudes in Human-Computer Interaction in general. Adams and Sasse (1999) used this approach to identify factors that affect employees' perceptions of corporate security policies, and Weirich and Sasse (2001) modelled employee decision-making on compliance with password security policies.

For the study reported in this chapter, only the sections on USB stick policies and tasks and situations surrounding their usage were analyzed. We coded the interviews using axial coding (the first stage of Grounded Theory) to produce an inventory of the individual employee's cost and benefit associated with USB stick usage, and the cost and benefit for the organization. The data were coded by two researchers independently.

The range of roles performed by the interview subjects was relatively diverse, from security managers to part-time researchers, as was the range and frequency of security related comments they produced. There were also noticeable differences in USB usage between the various interview subjects. From the interviews, we were able to identify two main USB stick usage scenarios. These scenarios broadly corresponded to the type of organization for which the subject worked. We have focused on the first of these scenarios in which the USB stick is used as a transport medium for data. This scenario is described in detail below. The second

scenario, corresponding to the research organization, in which the USB stick is also used as a primary data storage device, will not be covered here.

The following scenario is more representative of the financial services organization. In this scenario, the USB stick is primarily used for temporary storage for transit between locations such as an employee visiting a client company to deliver a presentation. The data required to deliver the presentation would be copied from the company's computer system onto the USB stick and taken to the client's location. Any data which must be brought back to the home company can be copied from the client's system onto the USB stick and brought back by the employee.

The data in this case is always backed up, either on the home company's system or the client company. The data is never unique and so a loss of a security stick cannot constitute a long-term availability issue. While a short-term loss of availability can be detrimental — the cost is to the individual, with possible small collateral reputation loss for the parent company if the clients need to resend data, etc. — it is unlikely to have a significant impact on the company.

A far bigger concern for the security manager in this scenario are the potential confidentiality issues resulting from company data being transported through unsecure locations while in transit to and from the client. If the USB stick were to be lost or stolen at this time, while containing unencrypted data, then the cost in terms of reputation and lost business would be to the company itself rather than the individual. While the company can punish the individual internally, it cannot recoup its losses by doing so. This scenario encourages the security manager to take a 'confidentiality first' approach when designing the USB control policy. We opted to focus on this scenario when describing our individual and organizational costs as it provided a relatively simple set of actions that encompassed the key points.

At this point we created a list of the actions required to complete the task in the scenario. This then was converted into a set of tables detailing the task at each stage, the cost to the individual, the cost to the organization, a possible failure mode at that juncture, and the cost to each of that failure. Appendix A (available at http://weis2008.econinfosec.org) contains the results of the empirical study, in tabulated form.

The data obtained in our empirical study, which has not been explored in this chapter, will be considered in future work.

4 The Conceptual Model

The empirical study discussed in Section 3 has presented ways in which USB sticks are used in two large organizations. In particular, this study shows that certain classes of events and risks arise during the course of the life-histories of USB sticks and their owners. This information provides a rich corpus that we can use to make modelling decisions. Accordingly, we have embodied these classes of

events and risks within the process model we now present. More specifically, we take, as the primary input to our model, the data obtained from the financial services organization.

For simplicity, we consider the organization of interest to consist in the collection of its individuals. Thus we can capture the behaviour of the organization, at this rather crude level of abstraction, by capturing the behaviour of a typical individual.

The purpose of our model is to embody the behaviour of our intended macroeconomics-inspired model of the relationship between the confidentiality and availability of information owned by an organization that uses USB memory sticks to support its operations. In this model, the instrument that is available to the organization is investment in information security. For the purposes of this study, we identify the following three types of investment:

- *Training* — individuals are trained to understand and work within the organization's information security policies;
- *IT Support* — the organization provides specialist IT personnel to help individuals resolve problems;
- *Monitoring* — the organization monitors the behaviour of the individuals with respect to its information security policies.

Our focus of attention for this model concerns the use of encryption of data held on USB memory sticks.

For each type of investment, we consider the idea of a *transfer function,* which associates to a given level of investment a certain parameter that is used to calculate the effect of a given level of investment. In the cases of *Training* and *IT Support*, the transfer function returns a value in the real interval [0,1]; in the case of *Monitoring*, the transfer function returns a (real) time interval. There are many reasonable choices for these functions, and we take simple exemplars, chosen primarily for their shape, on the presumption that more investment will generally increase the business proficiency and efficacy of the matter of interest, and they are guided by the following considerations:

- Whether they are monotonic increasing/decreasing;
- What limits they tend to;
- The presence of threshold effects for investment; and
- Algebraic simplicity.

We do not claim anything else for these particular functions — we do not know a priori what these functions ought to be, and so we leave that as an open question for further investigation. We consider them in turn.

First, the Training transfer function: The idea is that this transfer function takes the portion of the overall security investment budget allocated for training and specifies the probability of the individual making support calls. As the budget for training increases, the individual becomes more proficient and needs to make fewer and fewer support calls. We assume, however, that there is always a background

need to make some support calls, for example, having to do with aligning the USB encryption with organizational systems configurations. Thus the transfer function has output in [0,1] and is monotonically decreasing with increasing training budget. We further assume that a minimal amount of training is needed before there is any reduction in the probability of an individual making a support call. The form we have chosen for this function, where *inv* is the investment variable, is:

$$trainingTF\left(inv\right) = \left(b - c\right)\left(\min\left(1, a / inv\right)\right) + c\,,$$

illustrated in Figure 1; the parameters *a*, *b*, and *c* are defined as follows:

- *a* = minimum training investment threshold: The amount of investment needed before there is any effect on training and reduction on the probability of needing support;
- *b* = maximum probability of needing support: This value is attained when no training is given at all;
- *c* = minimum probability of needing support: We assume that there is a baseline, underlying need for IT support, no matter how trained the employees are. Clearly, we require $b \geq c$.

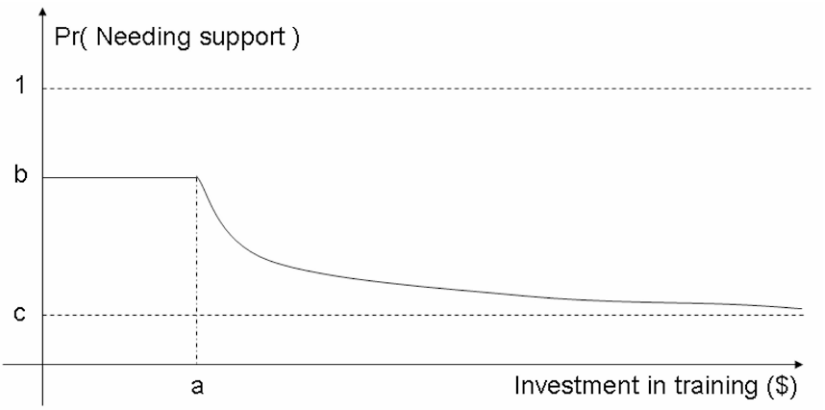

Figure 1. The 'Training' Transfer Function

Second, the IT Support transfer function: The idea here is that as security investment in IT support increases, the probability of a successful interaction with support also increases. The transfer function shows how this investment affects this probability, and is this time monotonically increasing. Just as for training, there is a minimum amount of investment required before any benefit is realised. The form we have chosen for this function is:

$$ITsupportTF\left(inv\right) = \max\left(0, b\left(1 - a / inv\right)\right),$$

illustrated in Figure 2; the parameters *a* and *b* are defined as follows:

- a = minimum IT support threshold: The minimum amount of investment required before there is any effect on the probability of the success of IT support;
- b = maximum probability of successful support: This is naturally a limiting value, which we assume can be achieved arbitrarily closely.

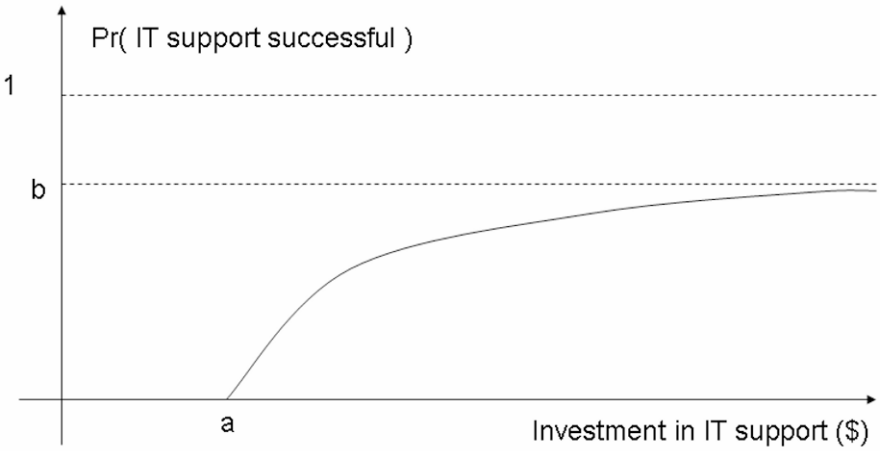

Figure 2. The 'IT Support' Transfer Function

Finally, the Compliance Monitoring transfer function: The idea here is that as security investment in compliance monitoring increases, this leads to an effective increase in the frequency with which compliance checks are made, so potentially improving the effectiveness of monitoring. Consequently, the time interval between checks will decrease. The transfer function specifying the time interval should therefore monotonically decrease as budgeted investment increases — the form of this function is conveniently chosen to be:

$$monitoringTF\left(inv\right) = \left(b - c\right)\left(\min\left(1, a / inv\right)\right) + c\,,$$

illustrated in Figure 3. The parameters a, b, and c are defined as follows:

- a = minimum monitoring investment threshold: The minimum amount of investment required before there is any reduction on the time interval between monitoring checks;
- b = maximum time interval between monitoring checks: A notional maximum amount of time between checks — in practice, this can simply be a very large number;
- c = minimum time interval between checks: It is assumed that each check must take some amount of time to complete — thus the time interval *between* these checks cannot be less than this. Clearly, we require $b \geq c$.

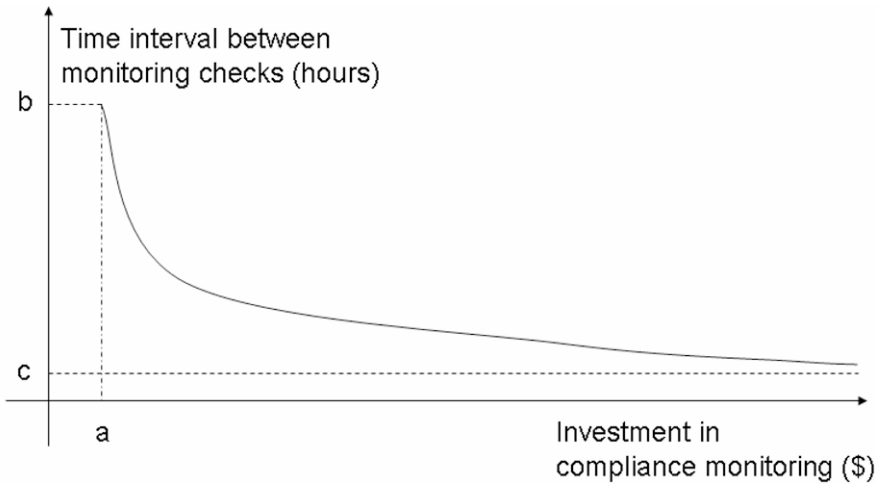

Figure 3. The 'Compliance Monitoring' Transfer Function

The transfer functions are used to determine the probability that a typical individual will employ encryption, in the manner intended by the security policy, when using a USB memory stick. Note that we are not in a position to give an analytic definition of this probability. Rather, this is the point at which we appeal to our empirical data and the simulations provided by our model (the code is given in Appendix B, available at http://weis2008.econinfosec.org). A key component of the model is the *individual's scoring function*,

EQ *indScore*:*R*\s\up5(4)→*R*\,

where *R* denotes the reals, expressing an individual's cost–benefit over the following four indicators:

- Successful data transfers (*trf*) — successful transfer of data is treated as a proxy for an individual's productivity;
- Embarrassments (*emb*) — events which damage the reputation of the individual, such as inability to recall a password in the presence of a customer;
- Reprimands (*ding*) — management may reprimand individuals for failing to comply with policy, and repeated reprimands may lead to serious sanctions;
- Negative experiences with IT Support (*nsup*) — interactions with IT Support may be unsatisfactory, and may fail to solve an individual's problem.

For the present study, we take the scoring function to be given by

$$indScore(trf,emb,ding,nsup) = dtSF(trf) + eSF(emb) + dSF(ding) + nsSF(nsup),$$

where *dtSF*, *eSF*, *dSF*, and *nsSF* are chosen functions that capture the dependency of the overall score on the evident components. Note that the scoring functions *eSF*, *dSF*, and *nsSF* are all negative-valued and decreasing because embarrassments,

reprimands, and negative IT Support experiences all have a negative impact on an individual's assessment of the cost-benefit trade-off of security activities.

As usual, there are many reasonable choices for these functions, and we take simple exemplars. In all cases, the specific functions used depend on some specific 'calibration parameters'. Rather than consider these parameters in detail, we explain here just the general form of the functions.

First, the scoring function for successful data transfers, illustrated in Figure 4, captures the existence of a limit on the maximum possible reward to the individual, no matter how high his productivity:

$$dtSF\left(trf\right) = a\left(1 - \frac{b}{trf + b}\right),$$

where $a,b>0$ are calibration parameters.

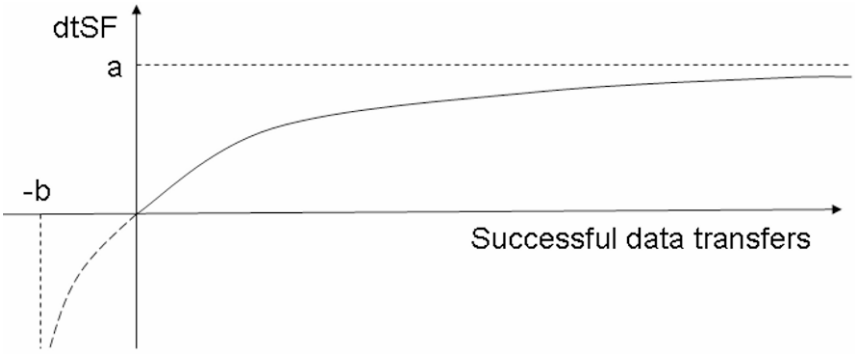

Figure 4. Individual Scoring Function for Successful Data Transfers

Personal embarrassments reduce the individual's score, so the scoring function *eSF*, illustrated in Figure 5, is negative decreasing; we assume that costs of embarrassments accumulate unboundedly:

$$eSF(emb) = -a(emb),$$

where $a>0$ is a calibration parameter.

Reprimands from management also reduce an individual's score, and the greater the number of reprimands, the smaller the effect of subsequent reprimands. The function *dSF*, illustrated in Section 6, has the following form:

$$dSF\left(ding\right) = a\left(\frac{b}{ding + b} - 1\right),$$

where $a,b>0$ are calibration parameters.

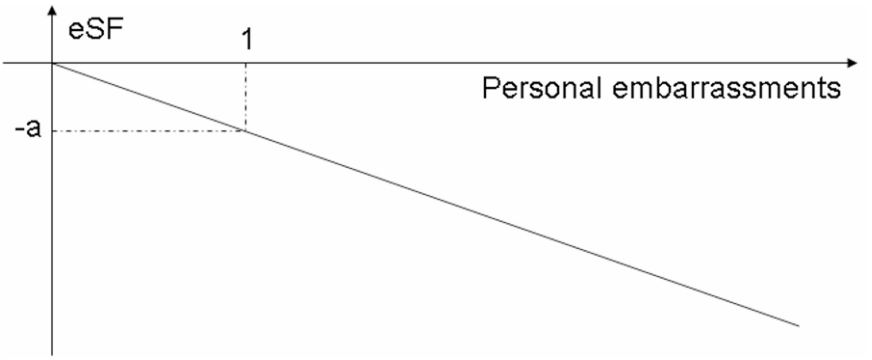

Figure 5. Individual Scoring Function for Personal Embarrassments

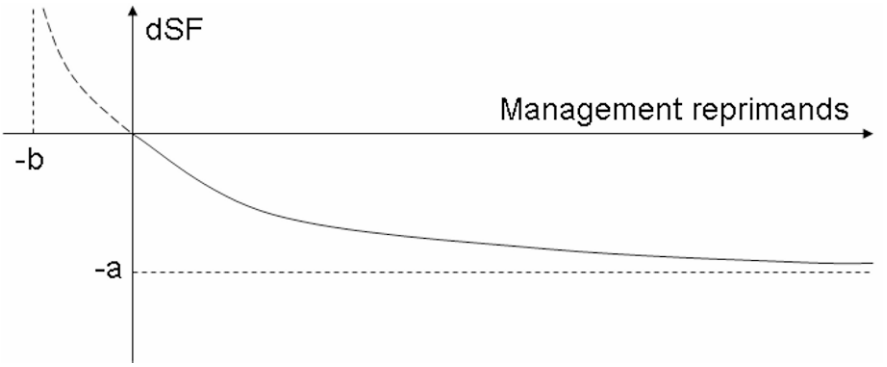

Figure 6. Individual Scoring Function for Management Reprimands

Finally, we consider the function *nsSF*, illustrated in Figure 7. Here we assume that the user's response to his failing to receive adequate support deteriorates as he experiences more such failures. We expect that it eventually overrides other factors, representing the encryption technology's becoming unusable and being given up. We take

$$nsSF\left(n\,sup\right) = -a\left(n\,sup^2\right),$$

with a calibration parameter a>0.

The typical individual's probability of using encryption is now obtained as follows:

- By using the above transfer and scoring functions, the model essentially becomes a function with a number of input parameters that maps over security investment, then security budget proportions, then probability of encryption, resulting in an overall numerical score as output. Informally,

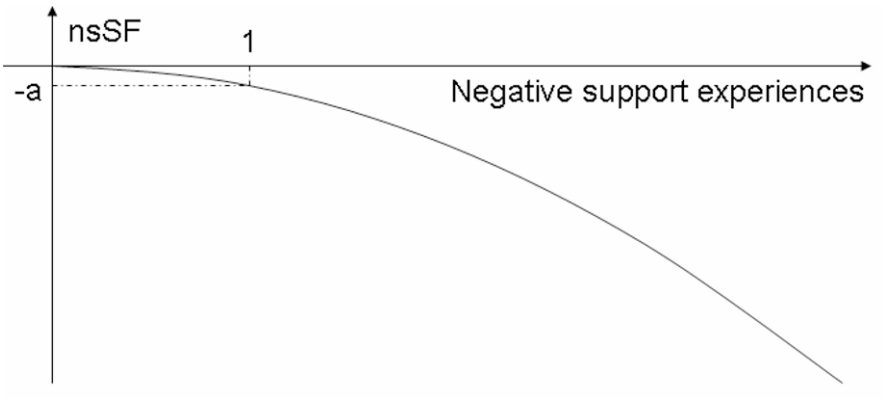

Figure 7. Individual Scoring Function for Support Failures

$$\text{model} : \text{security-investment} \rightarrow \text{security-budget-proportions} \rightarrow \text{probability-of-encryption} \rightarrow \text{score}.$$

Intuitively, this function represents the typical individual's score given all these input parameters. We also assume, however, that the typical individual responds rationally to the organizational environment (as determined by the security investment and the security budget proportions) by choosing how frequently he uses encryption, so as to maximize his perceived overall score. This rational maximization of benefit by the typical individual is therefore the basis for choosing the encryption probability;

- Mathematically speaking, our procedure for computing the probability p of encryption is to take $p \in [0,1]$ such that p is the (unique) value that maximizes the overall score as a function of security investment and security budget proportions:

$$\sup \{model(sec)(sec{-}budget)(p) \in R \mid p \in [0,1]\}$$

where $sec \in sec{-}range$ and $sec{-}range$ is a subset of R, representing the range of security investments to be investigated and where $sec{-}budget$ ranges over the budgetary splits we could make (e.g., IT support, etc.). Technically, this function might have several optima as p ranges over [0,1]; that is unlikely since the transfer and scoring functions are clearly monotonic (and also concave/convex) and we assume that they are sufficiently smooth for there to be a unique choice maximizing the score;

- This function is expressed in terms of an executable discrete event model involving stochastically generated events (see Section 5). Therefore, the numerical answers that we obtain are generally approximate. In effect, the computation we are making involves fixing discrete values for the security investment, the security budget proportions and then performing a range of experiments ranging over discrete values for the probability of encryption. Each of these

experimental variations are then performed a large number of times in order to obtain statistically valid outcomes from which we choose the probability value that maximizes the score. Intuitively, the multiple runs performed for each of the choices taken represents finding the average score over our typical population (we assume, for now, a homogeneous population).

The probability of using encryption has direct consequences for the utility function that derives from the model. The calculation of this function is explained in Section 6.

5 An Executable Model

The conceptual model described in the previous section is reified using our modelling tool, Demos2k (Demos2k; Birtwistle 1979], which executes discrete event models of systems of resources and processes. Demos2k has rigorous mathematical semantics (Birtwistle and Tofts 1993; 1994; 1998; 2001a; 2001b) based on process algebra (Milner 1983; 1989; Pym and Tofts 2006; 2007), which can be understood in both asynchronous and synchronous terms. Our modelling technique is to deploy the discrete mathematical tools of resource semantics (Pym 2002; Pym and Tofts 2006; 2007), process algebra (Milner 1989; Pym and Tofts 2006; 2007), and probability theory/stochastic processes (Demos2k; Tofts 1994] in the style of classical applied mathematics (see Yearworth et al. (2006) for another example of the approach); that is, we identify levels of abstraction that are appropriate to the questions of interest, and avoid representing irrelevant detail.

We model the life-history of the composite entity 'a typical individual together with his current USB stick' to illustrate how various forms of risk are encountered within a given amount of time. By modelling these risk encounters explicitly, we can obtain a better quantitative picture of how the risks identified are naturally distributed. Modelling this composite entity (i.e., the 'user') allows us to ignore aspects of an individual's own life that do not involve any dealings with the USB stick.

For there to be any risk to confidentiality or availability, we need to introduce some particular sources of hazard. For this investigation, there are two principal components contributing to the hazards that arise: the user's physical location and the categories of people with whom the user intentionally or unintentionally shares data. For the purposes of this model, we broadly categorize the people we share data with as follows: whether they are a colleague or business partner who might legitimately share the information (i.e., a 'Friend'), or someone who will actively misuse the information gained to somehow harm the organization or the user (i.e. a 'Foe'), or, finally, someone who appears to the user as a Friend but *in actual fact* acts like a Foe (i.e., a 'Traitor'). Both of these aspects — location and categories of people we share data with — are explicitly represented in the model.

The outcome of running the model will be values of various performance indicators gathered as a part of simulating the life-histories:

- Number of successful data transfers to/from the USB device: This is used as a straightforward proxy for productivity — we assume that using a USB stick to transfer data has business benefit;
- Total number of exposures: Occasions on which information was transferred to either a Foe or a Traitor;
- Total number of 'reveals': A 'reveal' is less significant than an exposure and arises when a colleague or business partner (i.e., a Friend) is given information that they did not have a right to see. Because they are Friends, they are not expected to use that information to cause harm to the organization or the user. One way in which this can arise is via 'accidental archiving' — information that was unintentionally made available alongside other information that was intended to be shared.

Various other indicators are also gathered as output from each run; these have already been discussed in Section 4.

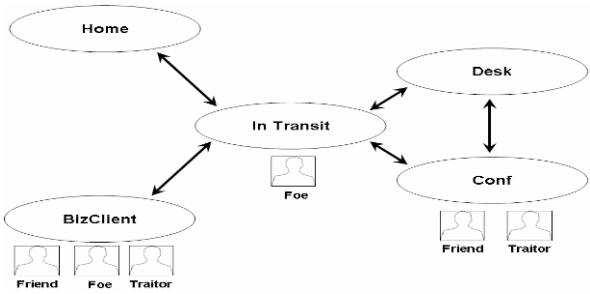

Figure 8. Locations and Roles

The model comprises three main concurrent processes: *lifeUSB*, *movement*, and *measure*:

- *lifeUSB*: This process captures the activities of the 'individual plus his USB stick'. The user essentially interacts with different kinds of people in different locations, and indicators are accumulated as a result. Particular events involving the USB stick, such as *add/modify*, *write*, *delete*, etc., are randomly selected according to (discrete) probability distributions, conditional upon current location. As a result of these actions and interactions, we use a combination of time penalties and indicators to capture and account for the risks encountered.
- *movement*: This process concurrently and independently moves the user from location to location, spending some time in each place. The different locations we use are:
- *Home*: The user's personal home;
  - *Desk*: The main place of (solitary) work for the user;
  - *Conf*: This is where business meetings with Friends (and, potentially, Traitors) occur;

- *BizClient*: Business meetings/workshops/conferences with business partners or other actors (i.e., principally Friends, but with some potential for talking to Traitors and Foes);
- *InTransit*: This represents intermediate locations (e.g., on a plane, in a hotel, in a car) between main locations.

  Each location naturally has its own associated risks and opportunities for interaction. The transitions between locations follow the graph presented in Figure 8. Note that we assume that the user can move directly between the workplace locations Desk and Conf without going via the riskier InTransit location. Future locations for the user are chosen according to a location-dependent probability distribution, as well as the period of time they spend there;
- *measure*: A bookkeeping process that samples the various indicators gathered on a regular basis throughout each run.

The Demos2k code for the model we described is given in Appendix B, available at http://weis2008.econinfosec.org.

6 The Experimental Space

Now that we have our executable model, we can use it to explore how the level of security investment by an organization is connected to particular levels of availability and confidentiality, as modulated and affected by changes in typical employee behaviour, vis-à-vis his use of USB memory sticks. The organization's choices of amount and balance of security investment affect the usage of encryption on USB sticks by typical employees. This usage results in levels of information availability and confidentiality loss, which translate into business value for the organization.

Our experiments, performed using Demos2k (Demos2k) and its DXM experiment manager (Monahan 2008), varied the following numerical instruments:

- Security Investment: This indicates the level of yearly investment per individual in security related cost. The range we have explored is: 20, 50, 100, 200, 500;
- Budgetary Proportions: Although we have three areas in which to invest — *training, IT support* and *monitoring* — we have specified a fixed value of training, since it is a one-off cost. So we have investigated the trade-off between investment in IT support on the one hand, and monitoring on the other. In practice, we have chosen to investigate 3 values of support proportion: 0.25, 0.5 and 0.75[30].

[30] A support proportion of 0.25 means that 1/4 of the total security investment goes towards IT support and the remainder goes towards monitoring.

Each of these 15 (3×5) sample points represents a particular experimental variation. Following the approach to obtaining the individual's probability of using encryption, explained in § 4, within each of these variations we then need to range over *Pr*(*Enc*), the probability of encryption, (from 0.1 to 0.9 in steps of 0.2) and finally run each of these 300 times to obtain results of some statistical value.

For simplicity of presentation in this chapter, we have had to restrict the number of experimental simulations, and so we have adopted a coarse-grain 'sampling' strategy to choose parameters. We plan to conduct a more thorough and systematic experimental investigation based on empirical evidence to support the form of the transfer and scoring functions; where that is not possible, we hope to perform a systematic investigation of the space of parameters. The objective of such an investigation is to provide detailed guidance for conditioning economic models of the kind we have discussed.

6.1 Exploratory Fit of Additional Calibration Parameters

The transfer and scoring functions given are each dependent upon a number of numerical parameters. At this stage, it has not been possible to find obvious choices for these parameters — there are no easy and obvious sources of data, and there are no 'natural scales' that we could obviously exploit in order to make considered and easily justified choices. Further empirical study and experimental work will be necessary to address this issue.

Instead, we have taken the pragmatic decision to make choices of these parameters that illustrate a range of behaviour. To do this, we have conducted a series of exploratory (ad hoc) searches through the space of additional calibration parameters, helping to locate values of these parameters that yield useful observable output. We cannot claim therefore that this study has given definitive or canonical results. We instead claim that there is evidence here for examining the connections between these concerns in greater depth.

6.2 Some Confirmation of Expected Behaviour

As investment in monitoring and IT Support increased, we expected to see greater use of encryption; that was observed.

We expected to see a variation in the effectiveness of that investment as the proportion spent on IT Support vs. Monitoring was varied. As illustrated by the results below, we did not observe any such effect: the influence of a given level of investment is roughly the same for different proportions. We expected to be able to see a gradual increase in the use of encryption as investment increased, but the results show a fairly sharp transition from probability of encryption of 0.1 to 0.9 between investment values of 100 and 200. (Examining the data in more detail than

shown here emphasizes this effect. The individual's optimal choice of probability (as computed from the experimental results) is always at one of the extremes, and never at a middle value.) We also expected that, above and below certain limits, there would be little extra effect from further increasing or reducing the investment level: this is not contradicted by the model (it is mildly confirmed).

6.3 Results

In Section 4, we described how to extract information about our estimate for *Pr(Enc)* for a given level of security investment and budgetary proportions, based upon the individual's scoring function. Intuitively, this value is the one that produces the maximum value of this scoring function at that investment level.

The table below gives the value of *Pr(Enc)*, for the budgetary proportion dedicated to IT support versus security investment:

Table 1. Value of *Pr(Enc)*

	20	**50**	**100**	**200**	**500**
0.25	0.1	0.1	0.1	0.9	0.9
0.5	0.1	0.1	0.1	0.9	0.9
0.75	0.1	0.1	0.1	0.9	0.9

This table shows that, for security investment of 100 and below, the user's best choice is *Pr(Enc)* = 0.1; that is, rarely to use encryption. For security investment of 200 and above, the user's best choice is *Pr(Enc)* = 0.9; that is, nearly always to use encryption. (We did not consider *Pr(Enc)* of 0 or 1 because such utterly consistent user behaviour is rare.)

Next we tabulate the observed values of the availability measure and of the confidentiality measure over the 15 sample points, with the user's *Pr(Enc)* fixed at the corresponding value shown in the table above.

The availability measure is chosen to be the average number of successful data transfers per year carried out by the user. This is under the assumption that the purpose of the USB stick is to enable the user to transfer data on behalf of the organization.

Table 2. Availability Measure

	20	**50**	**100**	**200**	**500**
0.25	165.093316	164.0433177	165.106651	161.2066513	161.1899847
0.5	163.453318	163.5266511	165.5766509	162.6299845	161.453318
0.75	164.729983	165.6333176	164.2733177	161.2266513	161.6966513

The confidentiality measure we use is a linear combination of the average number of events when confidential data is exposed and the average amount of confidential data exposed, both per year.

Table 3. Confidentiality Measure

	20	50	100	200	500
0.25	10.02999905	8.26666588	9.326665779	5.85666611	6.626666036
0.5	8.176665889	7.876665917	9.123332465	6.106666086	6.886666012
0.75	9.519999094	7.966665909	8.569999185	6.449999386	5.486666145

We can observe that there is a substantial change in both the organization's availability and confidentiality measures as the user's probability of using encryption, *Pr(Enc)*, changes from 0.1 to 0.9.

The results are all obtained as averages over 300 independent runs. These values conservatively have a standard error of less than 10% of the values in the table. Given the number of runs required, it seems that the standard error might be halved by performing 1200 runs.

All of these results are preliminary. Further, and quite extensive, experimental work will be required to obtain adequate confidence interval estimates for the numbers quoted above.

6.4 A Utility Function

We have discussed, in Section 2, a utility function approach to understanding the trade-offs between availability and confidentiality. We suggest that the simplest utility function it seems reasonable to postulate is one of the form

$$U(C, A) = \alpha(A - \beta C),$$

where α and β are parameters, which captures a simple ratio between confidentiality and availability.

Below are some tabulations of values for this function for different values of α,β, based upon the tables of availability and confidentiality numbers presented above. Exploring parameters of the utility function, illustrated in the tables below, we see that for values of β=10 or 3, as spending on support and monitoring increases, the gain from increased confidentiality clearly outweighs the consequent loss of availability. β=0.1 results in the loss in availability as spending increases outweighing the gain in confidentiality. Values of β in the region of 1 didn't give us useful results for utility, because statistical variation in experimental results swamps the difference between availability and confidentiality components of utility.

Table 4. Utility Function for α=1.164, β=10.000

	20	50	100	200	500
0.25	75.44987339	94.76066264	83.65550334	119.52117	110.5353456
0.5	95.12164829	98.70045181	86.57055909	118.2674243	107.814368
0.75	80.96557825	100.1055786	91.49626593	112.6352725	124.4002981

Table 5. Utility Function for α=0.714, β=3.000

	20	50	100	200	500
0.25	96.33782579	99.36347244	97.85302782	102.4985371	100.8382374
0.5	99.13512178	99.82968844	98.62371136	102.979025	100.4695461
0.75	97.17035435	101.1403263	98.87822724	101.2426083	103.6496

Table 6. Utility Function for α=0.615, β=0.100

	20	50	100	200	500
0.25	100.9077518	100.3704883	100.9592029	98.77427677	98.71667626
0.5	100.013201	100.0767461	101.2607345	99.63418527	98.86262497
0.75	100.7156816	101.3667113	100.4932739	98.75008864	99.09835674

7 Conclusions and Directions

We have reported a preliminary study. We have postulated an economic model that is suitable for capturing the utility of trade-offs between investments against confidentiality and availability in the context of the use of USB memory sticks in a financial services company. Building on empirically obtained data and on informed observations concerning policy and technology, we have used a process model to demonstrate that the hypothesized trade-off between confidentiality and availability does indeed exist, so providing evidence for the validity of the model, and to investigate the behaviour of a simple version of this model, giving good evidence to support the approach and motivate further study. We have established that individuals make cost–benefit decisions from their own (economic) perspective; we suggest organizations must understand that when making investment decisions.

The following is a brief list of possible research directions:

- Further exploration of our experimental space, with substantial statistical analyses to inform the detailed formulation of economics models of the kind we have discussed;
- Mathematical and computational studies of the properties of these models;

- An investigation of game-theoretic approaches to the utility of the allocation security investment resources against competing priorities such as confidentiality and availability;
- More basic empirical studies of the kind we have described; for example, more studies of portable data storage media, or studies of network access control policies;
- Developments of our process modelling tool better to handle the structure of distributed systems.

The work reported here is the result of a highly interdisciplinary study. Such an approach seems to us to be necessary to make progress in this area.

Acknowledgments

We are grateful to the many members of staff of HP Labs and Merrill Lynch who generously gave their time to take part in our empirical studies. We also thank Jean Paul Degabriele for his advice on the early stages of this work.

References

Anderson. R., and Moore, T. "The Economics of Information Security," *Science* (314), 2006, pp. 610–613. Extended version available at http://www.cl.cam.ac.uk/~rja14/Papers/toulouse-summary.pdf.

Anderson, R. "Why Information Security Is Hard: An Economic Perspective," in *Proceedings 17th Annual Computer Security Applications Conference*, 2001.

Adams, A.L., and Sasse, M.A. "Users Are Not the Enemy: Why Users Compromise Security Mechanisms and How to Take Remedial Measures," *Communications of the ACM* (42:12), 1999, pp. 40–46.

Barro, R., and Gordon, D. "A Positive Theory of Monetary Policy in a Natural Rate Model," *Journal of Political Economy* (91), 1983, pp. 589–610.

Birtwistle, G. *Demos — discrete event modelling on Simula*. Macmillan, 1979.

Birtwistle, G., and Tofts, C. "An Operational Semantics of Process-Orientated Simulation Languages: Part I,", Demos. *Transactions of the Society for Computer Simulation* (10:4), 1993, pp. 299–333.

Birtwistle, G., and Tofts, C. "An Operational Semantics of Process-Orientated Simulation Languages: Part II," Demos. *Transactions of the Society for Computer Simulation* (11:4), 1994 pp. 303–336.

Birtwistle, G., and Tofts, C. "A Denotational Semantics for a Process-Based Simulation Language," *ACM ToMaCS* (8:3), 1998, pp. 281–305.

Birtwistle, G., and Tofts, C. "Getting Demos Models Right — Part I Practice," *Simulation Practice and Theory* (8:6-7), 2001, pp. 377–393.

Birtwistle, G., and Tofts, C. "Getting Demos Models Right — Part II ... and Theory," *Simulation Practice and Theory* (8:6-7), 2001, pp. 395–414.

Mathematica Documentation Center. http://reference.wolfram.com/mathematica/guide/Mathematica.html. 2008.

Clatworthy, M., Peel, D., and Pope, P. "Are Analysts' Loss Functions Asymmetric?" *Technical Report 005*, Lancaster University Management School, 2006.

Demos2k. http://www.demos2k.org.

Gordon, L.A., and Loeb, M.P. "The Economics of Information Security Investment," *ACM Transactions on Information and Systems Security* (5:4), 2002, pp. 438–457.

Gordon, L.A., and Loeb, M.P. *Managing Cybersecurity Resources: A Cost-Benefit Analysis*. McGraw Hill, 2006.

Milner, R. "Calculi for Synchrony and Asynchrony," *Theoretical Computer Science* (25:3), 1983, pp. 267–310.

Milner, R. *Communication and Concurrency*. Prentice-Hall, 1989.

Monahan, B. "DXM: Demos Experiments Manager," Forthcoming *HP Labs Technical Report*, 2008.

Nobay, R.A., and Peel, D.A. "Optimal Discretionary Monetary Policy in a Model of Asymmetric Bank Preferences," *The Economic Journal* (113:489), 2003. pp. 657–665.

Pym, D., and Tofts, C. "A Calculus and Logic of Resources and Processes," *Formal Aspects of Computing* (18:4), 2006, pp. 495–517, Erratum (with Collinson, M.) *Formal Aspects of Computing* (19) 2007, pp. 551–554.

Pym, D., and Tofts, C. "Systems Modelling via Resources and Processes: Philosophy, Calculus, Semantics, and Logic," in Cardelli, L., Fiore, M., and Winskel, G. (Eds), *Electronic Notes in Theoretical Computer Science (Computation, Meaning, and Logic: Articles dedicated to Gordon Plotkin)*, (107) 2007, pp. 545–587, Erratum (with Collinson, M.) *Formal Aspects of Computing* (19) 2007, pp. 551–554.

Pym, D.J. *The Semantics and Proof Theory of the Logic of Bunched Implications*, *Applied Logic Series* 26 Kluwer Academic Publishers, 2002. Errata and Remarks maintained at: http://www.cs.bath.ac.uk/~pym/BI-monograph-errata.pdf.

Ruge-Murcia, F.J. "The Inflation Bias When the Central Bank Targets the Natural Rate of Unemployment," *Technical Report 2001-22*, Département de Sciences Économique, Université de Montréal, 2001.

Ruge-Murcia, R.J. "Inflation Targeting under Asymmetric Preferences," *Journal of Money, Credit, and Banking* (35:5), 2003, pp. 763–785.

Strauss, A.L., and Corbine, J.M. *Basics of Qualitative Research: Grounded Theory Procedures and Techniques*. Newbury Park, CA: Sage, 1990.

Office Of Science and Technology. Foresight: Cyber Trust and Crime Prevention Project: Executive Summary. 2004.

Taylor, J.B. "Discretion versus Policy Rules in Practice," *Carnegie-Rochester Conference Series on Public Policy* (39), 1993, pp. 195–214.

Tofts, C. "Processes with Probability, Priority and Time," *Formal Aspects of Computing*, (6:5), 1994, pp. 536–564.

Varian, H. "A Bayesian Approach to Real Estate Management," in Feinberg, S.E. and Zellner, A. (Eds) *Studies in Bayesian Economics in Honour of L.J. Savage*, North Holland, 1974, pp. 195–208.

Weirich, D., and Sasse, M.A. "Pretty Good Persuasion: A first Step towards Effective Password Security for the Real World," in *Proceedings of the New Security Paradigms Workshop,* Cloudcroft, NM, ACM Press. September 2001, pp. 137–143.

Yearworth, M., Monahan, B., and Pym, D. "Predictive Modelling for Security Operations Economics," (extended abstract) in *Proc. I3P Workshop on the Economics of Securing the Information Infrastructure*, 2006. Proceedings at http://wesii.econinfosec.org/workshop/.

Zellner, A. "Bayesian Prediction and Estimation Using Asymmetric Loss Functions. *Journal of the American Statistical Association* (81), 1986, pp. 446–451.