FACULTAD DE CIENCIAS EMPRESARIALES Y ECONOMIA

Serie de documentos de trabajo del Departamento de Economía /
Department of Economics Working Papers Series

THE PROBLEM OF PREVENTION*

Jean- Pierre Benoit
Economics Area
London Buisness School
jpbenoit@london.edu

Juan Dubra
Departamento de Economía
Facultad de Ciencias Empresariales y Economía
Universidad de Montevideo
jdubra@um.edu.uy

Working paper UM_CEE_2011-11
http://www.um.edu.uy/cee/investigaciones/

The Problem of Prevention*

Jean-Pierre Benoît
Economics Area,
London Business School

Juan Dubra
Universidad de Montevideo

December, 2011

Abstract

Many disasters are foreshadowed by insufficient preventive care. In this paper, we argue that there is a true *problem of prevention,* in that insufficient care is often the result of rational calculations on the part of agents. We identify two factors that lead to dubious efforts in care. First, when objective risks of a disaster are poorly understood, positive experiences may lead to an underestimation of these risks and a corresponding underinvestment in prevention. Second, redundancies designed for safety may lead agents to reduce their care, resulting in a decrease in safety under certain conditions. We also analyze the use of checklists in accident prevention.

Keywords: Prevention, Accidents, Volunteer's Dilemma, Learning, Checklists.
Journal of Economic Literature Classification Numbers: D81, D82, D83

A remarkable number of disasters and near-disasters, from the nuclear mishap at Three Mile Island,[1] to the Union Carbide plant tragedy in Bhopal,[2] to the Challenger disaster,[3]

*We thank Martín Besfamille, Federico Echenique, Emeric Henry, Ennio Staccetti and Federico Weinschelbaum for their comments. Benoît acknowleges the support of the C.V. Starr Center at NYU, and Dubra thanks the support from ANII through the Fondo Clemente Estable.

[1]In March 1979, there was a partial meltdown of the reactor core of the Three Mile Island Unit 2 nuclear power plant.

[2]In December 1984, methal isocyanate gas was released at the Union Carbide chemical plant in Bhopal, India, resulting in thousands of deaths and hundreds of thousands of injuries.

[3]The American space shuttle Challenger exploded shortly after takeoff on January 28, 1986.

to Hurricane Katrina[4] have been preceded by a woefully inadequate level of preventative care, making these adverse events seem not so much manifestations of poor luck, as all but inevitable occurrences. Indeed, the phrase "an accident waiting to happen" has become somewhat of a cliché in post-event reporting. In a similar vein, a study by the Institute of Medicine (2000) concluded that each year over 44,000 people die in US hospitals from preventable medical errors. In the banking industry, huge losses have resulted from a succession of rogue traders despite safeguards put into place with each episode. In this paper, we argue that there is a true *problem of prevention,* in that many accidents are waiting to happen as the result of rational calculations on the part of agents. We identify two factors that lead to dubious efforts in care.

1. When objective risks of a disaster are poorly understood, positive experiences may lead to an underestimation of these risks and a corresponding underinvestment in prevention.

2. Redundancies designed for safety may lead agents to reduce their care, resulting in an overall decrease in safety under certain conditions.

We use a simple model of accident prevention that captures these two features. Recent work has argued that the use of checklists may significantly reduce the likelihood of accidents in health care and other industries (see Gawande (2010) for an extended discussion) — we apply our model to the analysis of checklists.

Much of the writing on accidents comes from sociologists and psychologists. Vaughan (1996) has written an in-depth study of the Challenger accident in which she faults the culture of organizations, in general, and of NASA, in particular; Perrow (1999) has written about the danger of tightly coupled complex systems, such as Three Mile Island. Downer (2011*b*) argues that there is a category of *epistemic accidents* which result from flawed theories and judgements. Sagan (2004) and Downer (2011*a*) highlight some of the same issues that we discuss, among other things, but argue informally. We will return to this literature, and to the relevant economics literature, at various points in the paper.

At times, agents are simply poorly motivated – for instance, they may not see the full cost of damages, or they may discount too sharply – or simply make mistakes (Reason (1990

[4]Hurricane Katrina struck southeast Louisiana on August 29th. Considerable damage was caused, including the flooding of 80% of New Orleans. There were over 1500 deaths as a result.

studies various types of errors to which humans are prone). While these can be important factors, our interest, is in difficulties that remain even when actors are well motivated and well trained to avoid mistakes.

1 The Model

To fix our ideas, consider a machine with one critical part that may become defective and fail in any period, with some given unknown probability. In each period, prior to running the machine the part can be tested by several agents independently and, if found defective, costlessly repaired. The test itself, however, is costly and imperfect – at higher costs the test is more likely to detect a defect. In addition, an automated device may perform a test. We can think of a defective part as an *event*, which turns into an *accident* if and only if it is not detected. With this story in mind, consider the following simple model.

There are $k \geq 2$ agents, an automated device, and nature. In each period $t = 0, 1, 2...$, nature chooses $y \in \{e, n\}$ (an event occurs or no event occurs) according to some probability $\Pr(y = e) = \hat{\theta} \in (0, 1)$. The parameter $\hat{\theta}$ is unknown, and every agent has the same beliefs about $\hat{\theta}$. Given a probability distribution q over $[0, 1]$, the subjective probability of an event is denoted $\overline{\theta}_q = \int \theta dq(\theta)$.

In every period, each agent chooses an investment in care $c \in S \subseteq \mathbf{R}$, for some closed interval S. The choice of care is private information. If an agent invests c in care, he or she fails to detect (and fix) an event that has occurred with probability $p(c)$. The function p is twice continuously differentiable with $p' < 0$, $p'' \geq 0$. In addition, there is an automated device which may detect (and fix) an event. The automated device fails with probability p_a. An accident happens if and only if an event occurs and all agents and the automated device fail to detect it. If an event is detected, all agents are informed of it. An accident is so severe that it effectively ends the problem for the agents (allowing the agents to continue would not change our results).

Given a profile of effort choices $\mathbf{c} = (c_1, ..., c_k)$ in the current period, and an expected probability of event $\overline{\theta}_q$, the (subjective) probability of an accident, that is, the probability of an undetected event, is $\overline{\theta}_q p_a \pi_{i=1}^{k} p(c_i)$.[5] An accident causes a loss of D to an agent; the

[5] In our model, we assume that the probabilities that different parts of the sytem fail are independent of each other. Downer (2011a) argues that in many cases it is difficult to tell whether or not independence

payoff in any single period in which there is no accident is normalized to zero. Thus, the expected payoff of agent $i = 1, ..., k$ in the current period is $-\overline{\theta}_q p_a \pi_{i=1}^{k} p(c_i) D - c_i$.

We focus on Markov strategies. Specifically, we define the state to be agents' beliefs about $\hat{\theta}$ and consider strategies which depend only upon agents' current beliefs. Thus, we rule out an (arbitrary) dependence on time. Given an absence of strategic dependence across time, in every period agents seek to maximize their single period payoff.

We denote the above game by $G(k, q, p, p_a)$, where we suppress the dependence on S and D, which plays no role in what follows. We focus on symmetric equilibria, though we briefly discuss asymmetric equilibria in Section 2.

Theorem 1 *The game $G(k, q, p, p_a)$ has a unique symmetric equilibrium in Markov strategies.*

Proof. All proofs are in the appendix. ■

In the next two sections we perform comparative statics that elucidate some important aspects of the problem of prevention.

1.1 Good News Can Be Bad

First consider agents' beliefs about the inherent safety of their environment, that is, their beliefs about $\hat{\theta}$, the probability of an event. Scientific and other considerations yield a priori estimates which must then be updated in the light of experience. Some industries, such as the airline industry, have a long track record with both successes and failures, so that there is a good understanding of the pertinent probabilities — even when new engines and airplanes are developed, there is a good understanding of the ways in which these need to be tested.[6] Other enterprises, such as nuclear power plants and the space shuttle, involve relatively new technologies with limited experience. These spare histories make it very difficult to estimate the risks involved. In particular, unbroken strings of success make it difficult to assess the probability of a failure. As an example, the space shuttle Challenger had been preceded by twenty-four successful shuttle launches without a failure, and estimates of a catastrophic

holds.

[6]However, Downer (2011b) argues that some innovations in airplane design, such as the introduction of new composite material, may be poorly understood.

failure ranged from 1 in 100 to 1 in 100,000 (Feynman (1988))[7]. Similarly, prior to the incident at Three Mile Island there had not been a single accident at a commercial nuclear power plant, and the risks were poorly understood. The likelihood of some natural disasters is also difficult to assess.[8]

Any reasonable updating process has the feature that the more time that passes without an adverse incident, the lower the probability that is attached to one. This increasing optimism leads to a declining investment in precautionary care, and, potentially, to dangerously little care. In this respect, *good news can be bad*. Investigations into the meltdown at Three Mile Island and the space shuttle Challenger accident show that such optimistic underinvestment is precisely what took place.

With regard to Three Mile Island, the Kemeny Commission (1979) concluded that:

> "After many years of operation of nuclear power plants, with no evidence that any member of the general public has been hurt, the belief that nuclear power plants are sufficiently safe grew into a conviction. One must recognize this to understand why many key steps that could have prevented the accident at Three Mile Island were not taken. (p.9)."

With regard to the Challenger, as part of the investigating commission, Feynman (1988)[9] wrote:

> We have also found that certification criteria used in flight readiness reviews often develop a gradually decreasing strictness. The argument that the same risk was flown before without failure is often accepted as an argument for the safety of accepting it again. (p.220)
>
> The *Challenger* flight is an excellent example: there are several references to previous flights; the acceptance and success of these flights are taken as evidence of safety. (p223)
>
> The slow shift toward a decreasing safety factor can be seen in many [areas]. (p230)

[7] It should be noted, however, that the (management) estimate of 1 in $100,000$ is a little hard to rationalize. Benabou (2008) argues that the estimate is a result of "groupthink".

[8] Born and Viscusi (2006) argue that this is especially true of "blockbuster catastrophes".

[9] Feynman's appendix to the commission's report is reprinted in Feynman (1988).

Vaughan (1996) has termed this steady decline in standards the "normalization of deviance", though she ascribes a different mechanism to this decilne. This reduction in care also has similarities to what Sagan (2004) calls "overcompensation".

The following theorem formalizes this phenomenon. The theorem is also valid for the care taken by a single agent, $k = 1$, provided that $p'' < 0$. Given a prior q about θ, let q_n denote the Bayesian posterior beliefs after no event has happened.[10] Let $c(k, q, p, p_a)$ denote the individual level of care in the symmetric equilibrium of $G(k, q, p, p_a)$.

Theorem 2 *For any density q with support $[0, 1]$, the probability of an event under beliefs q_n is strictly smaller than under beliefs q. That is, $\overline{\theta}_{q_n} \equiv \int_0^1 \theta q_n(\theta)\, d\theta < \int_0^1 \theta q(\theta)\, d\theta = \overline{\theta}_q$. Moreover, the level of care taken in the symmetric equilibrium of $G(k, q, p, p_a)$ is increasing in $\overline{\theta}_q$, so that care falls after no event. That is, for any r such that $\overline{\theta}_r < \overline{\theta}_q$, $c(k, r, p, p_a) \leq c(k, q, p, p_a)$ with strict inequality if $c(k, q, p, p_a)$ is interior.*

Thus, a string of periods with no events leads to a reduced belief in the probability of an event and a decline in care.[11]

While this decline in the level of care is interesting in and of itself, the question remains as to whether or not it is proper; after all, it is the result of Bayesian updating. Absent an objective measure of the probability of an accident, the question cannot be definitively answered. Nonetheless, it is clear that both the Kemeny Commission and Feynman considered that a) at the time of the accident, agents were taking too little care, while b) initially they were taking the correct (or at least a reasonable) amount of care. Clearly, the pejorative term "deviance" indicates that Vaughan also considers the decline in care to be inappropriate.

To understand this attitude, let us think of those who set the care standards as, collectively, the principal, and those who actually take the care as the agents. We then have a principal-agent problem. Implicit in the situation is the presumption that the principal cannot simply take the care herself, and cannot adequately monitor the agents' actions. In a standard principal-agent problem, the "problem" arises from the fact that the principal and agent have different motivations. Here, we focus on a different problem, namely one that

[10] Recall that the agent observes an event even if the event does not turn into an accident. Under an alternate formlutation, the agent only observes events that turn into accidents. Adopting this alternate formulation would add an inferential complication.

[11] The net effect of these changes on the *subjective* probability of an accident depends on p/p', as detailed in Theorem 4.

arises from a discrepancy in the beliefs of the principal and the agent. We call this type of problem a *belief-based agency problem*.[12]

The basic idea in the present context is the following. The principal is an expert who conveys her information/beliefs to the agents, but (inevitably) does so imperfectly. While the principal may be able to convey her mean belief fairly accurately, she is unable to convey the breadth and depth of the information on which this belief is based. As a result, the agents react more to additional information than the principal deems optimal. Alternatively, the agents may believe that there is more idiosyncratic variation across, say, power plants, than the principal does, so that they overreact to the experience at their particular power plant.

Formally, suppose the principal has belief q', while the agents have belief q. Both q' and q are assumed to be represented by Beta distributions.[13] The Beta assumption is fairly unrestrictive, as any smooth unimodal density on $[0, 1]$ can be well approximated by a Beta density (Lee (1989)). Statisticians often posit a Beta distribution when studying the updating of Bernoulli priors.

First suppose that the distributions of the principal and the agents have the same mean, but that the agents' (common) distribution has a larger variance. Then, initially, the principal and the agents agree upon the optimal amount of care. However, as we show below, following any sequence of non-events, the agents are always more optimistic than the principal, and, hence, invest too little in care. In fact, we establish a more general result. To understand this result, first note that given two Beta distributions $B(a, b)$ and $B(d, e)$ with the same mean, it can be shown that $B(a, b)$ has a larger variance than $B(d, e)$ if and only if $a < d$ and $b < e$. We generalize this condition and say that the beliefs of an agent with prior $B(a, b)$ are *more disperse* than those of a principal with prior $B(d, e)$ if $a < d$ and $b < e$ (thus, we have removed the requirement of equal means).

If the agents' beliefs are more disperse than the principal's, then initially the agents may be either more or less optimistic, in terms of mean belief, than the principal. In

[12] Many public health campaigns surrounding lifestyle choices (such as the use of condoms, the decision to smoke, dietary choices) fall into this category – the government seeks to change behaviour by informing citizens of the risks involved, but typically finds that individuals' beliefs concerning these risks can only be influenced, not dictated.

[13] The Beta distribution $B(\alpha + 1, \beta + 1)$ has a density on $[0, 1]$ given by $f(x) = x^{\alpha}(1 - x)^{\beta} / \int u^{\alpha}(1 - u)^{\beta} du$, and a mean of $\frac{\beta+1}{\alpha+\beta+2}$. After an observation of an event, a prior $B(\alpha + 1, \beta + 1)$ is updated to $B(\alpha + 1, \beta + 2)$; after a non event it becomes $B(\alpha + 2, \beta + 1)$.

either case, as the following theorem indicates, following enough good news, the agents will be more optimistic than the principal, and underinvest relative to the principal's beliefs.[14] If the principal and agents begin with the same mean belief, then the agents will begin underinvesting following the first non-event. The theorem is also valid for the care taken by a single agent, provided that $p'' < 0$.

Let q_{n^t} be the (Bayesian) posterior of q following t observations of n, and recall that $\overline{\theta}_{q_{n^t}}$ is the estimated probability of an event based on the distribution q_{n^t}.

Theorem 3 *Suppose the beliefs of the agents, q, are distributed according to $B(a,b)$ and the beliefs of the principal, $\widetilde{q}$, are distributed according to $B(d,e)$. If the beliefs of the agent are more disperse than those of the principal, then enough non-events will make the agent more optimistic than the principal. Specifically, for all*

$$t > \max\left\{\frac{bd - ae}{e - b}, 0\right\} \equiv T^*$$

we have $\overline{\theta}_{q_{n^t}} < \overline{\theta}_{\widetilde{q}_{n^t}}$. For all $t \geq T^*$, $c(k, q_{n^t}, p, p_a) \leq c(k, \widetilde{q}_{n^t}, p, p_a)$ and if $c(k, \widetilde{q}_{n^t}, p, p_a)$ is interior, the inequality is strict.

An extreme case occurs when the principal's priors are so tight that he deems (essentially) no updating to be appropriate. This seems to have been the case with Three Mile Island and the Space Shuttle.

When the potential damage from an accident is very large, the optimal number of accidents is close to zero. For this reason, nuclear reactors are built so that a string of successes is the norm. Unfortunately, our results indicate that this success is to some extent self-defeating. The Kemeny Commission reached much the same conclusion about "overupdating" on the part of power plant operators. In its report it states:

> The Commission is convinced that this attitude [namely, the inference that nuclear plants are safe based on their positive record] must be changed to one that says nuclear power is *by its very nature* potentially dangerous, and, therefore, one must continually question whether the safeguards already in place are sufficient to prevent major accidents (emphasis added). (p9)

[14] In essence, the agent learns more from the positive signals than the principal does. The problem of which priors are subject to more learning has not, as far as we know, been studied in general.

In effect, the commission is imploring nuclear operators to ignore favorable experience pointing to the safety of nuclear plants. Some of Feynman's recommendations can similarly be interpreted as exhortations to downplay the significance of experience. However, it is difficult, if not impossible, to prevent agents from engaging in their own updating. Moreover, at least two factors exacerbate this difficulty. The first one is the possibility of idiosyncratic differences. Consider airplane pilots. It is only natural, though perhaps unfortunate, for a particular pilot without an adverse incident to think of himself or herself as particularly skilled, and to be correspondingly less wary than overall probabilities would recommend. Similarly, operators at nuclear power plants may well feel that general experience at plants does not account for the specific conditions at their specific plants. The second factor is the phenomenon that, when estimating probabilities, people tend to place undue weight on factors that they can readily recall (the so-called availability heuristic), chief among these being their personal experience.[15]

Theorem 3 affords another interpretation beyond the principal-agent one. Some industries, such as airplanes, are well understood, not only because of their long experience, but also because they are built "bottom up." In contrast with conventional aircraft, the space shuttle was built with a "top down" approach (Feynman (1988)), making it difficult to obtain a tight estimate of the safety of its novel technology. Let the priors $\widetilde{q}$ correspond to well-established and time-tested technologies, and the priors q correspond to new or innovative technologies for which less is known. With that reading, Theorem 3 tells us that innovative technologies are especially susceptible to good news being bad.

We turn now to some related literature.

Our model points to the interaction between learning and investment. As is well understood, for static problems in which the decision maker is an expected utility maximizer, it does not matter whether agents know the probability of an accident, or whether they merely have a distribution of probabilities. When the problem of prevention is repeated over time, however, learning and care-taking interact in non-trivial ways. Gollier (2002) has studied how the curvature (and higher derivatives) of the utility function of the decision maker affect the optimal initial level of care taken when the probability of the accident is unknown. In contrast, our main concern is the study of the evolution of beliefs and how this evolution

[15] Our results suggest a line of research into the optimal incentive schemes for belief-based agency problems. We do not pursue such an investigation in the present paper.

affects investment over time.

One of the main features of our model, that strings of successes lead to lower care, is reminiscent of the search literature when the distribution that generates wage offers is unknown. This literature has shown that as time goes by, a worker who keeps receiving bad offers becomes more pessimistic about his prospects of finding a decent paying job. He then reduces his reservation wage. The first papers to analyze the decline in reservation wages were, under different assumptions, Rothschild (1974) and Burdett and Vishwanath (1984). Dubra (2004) studies the consequences of this decline on the welfare of the decision-maker.

1.2 Redundancy Systems

A lifeguard must continually scan a pool or a beach for signs of swimmers in distress. Unfortunately, even highly trained lifeguards may fail to maintain the necessary vigilance.[16] Theorem 2 suggests that lifeguards who face few emergencies will be especially prone to lapses in vigilance. This finding is consistent with experimental work in psychology which shows that subjects engaged in vigilance tasks perform relatively poorly when the signal rate is low.[17]

While the meandering mind of a lifeguard may prove lethal, the danger posed pales in comparison to the potential harm from a nuclear or chemical plant. For this reason, these plants are designed so that the complacency of a single individual is not sufficient for a disaster to ensue. Consider the following description of an incident at a Union Carbide plant in Institute, West Virginia (Perrow (1999)):

> “[Dangerous] Aldicarb oxime... was transferred to a standby tank that was being pressed into service because of some other problems. Unfortunately, the operators did not know that this tank had a heating blanket and that it was set to come on as soon as it received product. Also unfortunately, they were not examining the appropriate temperature gauges because they thought there was no need to, and there may have been problems with these anyway because of the nature of the product in the tank. A couple of warning systems failed to activate, and the tank blew... . A few other failures took place...” (p. 358)

[16] A 2001 Jeff Ellis & Associates study conducted at 500 swimming pools found that only 9% of lifeguards spotted a submerged mannequin within 10 seconds (considered crucial), and only 43% within 30 seconds.

[17] See Davies and Parasuraman (1981) for a survey.

Note the number of elements that fell into place to produce this accident: a standby tank was being used *and* there was a heating blanket *and* it was set to come on *and* the operators did not check the temperature gauges *and* warning systems failed *and* the tank blew *and* ... still other things happened. Even with all these failures, there was no loss of life, partly because weather conditions were propitious.

Certainly, the large number of factors that must align in order to produce an accident at a chemical plant contributes to its safety.[18] More generally, consider a system with numerous safety features, all of which must fail for a disaster to result. If the features might fail with given independent probabilities, then the more features, the safer the system.[19] With fully automated features, the logic is unassailable. If humans are involved, however, features that are ostensibly independent may manifest a *strategic dependence*, resulting in an ambiguous relationship between reliability and the number of features.[20]

Returning to the Union Carbide case described above, the mere failure of the operators to check the temperature gauges was a long way from producing an accident.[21] But why did the operators fail to check the gauges? The immediate reason given is that "they thought there was no need to," but why did they feel no need to follow such an elementary safety precaution? In this section we suggest that at least part of the reason was that the operators knew that even with this lapse, an accident was unlikely, precisely because so many factors had to go awry in order to produce one. That is, the very redundancy features which enhanced the safety of the plant also reduced the incentive of agents to take care, thus

[18] Perrow (1999), however, emphasizes the *dynamic* danger of tightly coupled complex systems, such as chemical plants. When things start to go wrong in these systems, it is difficult for workers to understand exactly where the problem lies and how to remedy it on the fly. Thus, whereas we take a static view in our modelling, Perrow is concerned with dynamic difficulties. Nonetheless, Perrow concedes that the number of failures that must take place for an accident to occur, per se, provides a crucial measure of safety.

[19] Sagan (2004) points out that adding redundancies may be counterproductive if the failure of one part may itself cause the failure of another. This possibility is absent from our model.

[20] A recent related, but different, literature models situations of interdependent risks where the probability that one player suffers a loss depends on the efforts of other players (for instance, the success of one airline's anti-terrorism efforts is affected by the actions of other airlines with connecting luggage). In this literature the actions of agents are assumed to be contractible. For examples, see Heal and Kunreuther (2007) and Kunreuther and Heal (2003), and the references therein.

[21] Similar lapses in care have been noted at numeous other accident sites, including Three Mile Island.

limiting the degree of safety that could be achieved.[22] An estimation of the safety of the system that neglects this strategic slackening will badly miss the mark.

While these strategic reductions in individual care raise the probability of a disaster, increases in the number of people and improvements in automation, in and of themselves, lower this probability; the net effect is ambiguous. Importantly, under some reasonable conditions, the net effect of adding redundancy features is an increase in the probability of a disaster. The following theorem summarizes these findings. Again, the theorem remains valid for the care taken by a single agent, provided that $p'' < 0$.

Recall that $c(k, q, p, p_a)$ is the individual level of care in the symmetric equilibrium of $G(k, q, p, p_a)$, and let $P(k, q, p, p_a)$ be the equilibrium probability of an accident.

Theorem 4 *In the unique symmetric equilibrium of $G(k, q, p, p_a)$,*

i) c is decreasing in k and increasing in p_a.

ii) P may be increasing or decreasing in its arguments.

Consider $k' > k$ and $p'_a > p_a$ and suppose the equilibrium is interior

(i.e., there is a $\bar{c} \in S$ such that $-p(0)^{k-1} p'(0) > \frac{1}{\bar{\theta}_q p_a D} > -p(\bar{c})^{k-1} p'(\bar{c})$).

If $\frac{p}{p'}$ is strictly increasing, then $P(k', q, p, p_a) > P(k, q, p, p_a) > P(k, q, p, p'_a)$;

if $\frac{p}{p'}$ is strictly decreasing, then $P(k', q, p, p_a) < P(k, q, p, p_a) < P(k, q, p, p'_a)$.

Mathematically, an increase in p_a is equivalent to an increase in $\bar{\theta}_q$. Thus, Theorem 4 tells us that a system that is inherently unsafe, may have fewer accidents than a relatively safe system.[23] A moment's thought makes this contrary finding clear. Suppose there is a single agent who can either take no care or perfect care. That is, $S = \{0, 1\}$ and $p(0) = 1$, $p(1) = 0$. For small enough $\theta_q > 0$ it is optimal for the agent to take no care, resulting in an accident probability of θ_q, while for large θ_q it is optimal for the agent to take perfect care, resulting in an accident probability of 0.

Examples 2 and 3 below, and the example in Section 2, are special cases of a function $p(c) = (1 - ac)^b$, with $a, b > 0$, for which p/p' is increasing. An example for which p/p' is decreasing is $p(c) = a(1 + c)^{-\gamma}$, for $a, \gamma > 0$.

[22] Sagan (2004) and Downer (2011*a*) argue informally that redundancies may lead to decreasing care.

[23] Viscusi (1984) argues that child safety caps on aspirin led to a decrease in adult care. While he offers no theoretical argument on the net safety impact, his empirical analysis suggests that the decrease in care offset the benefit of the safety cap.

Although the statement of the theorem is in terms of whether p/p' is monotonically increasing or decreasing, even if p/p' is not monotone over the entire domain, the comparative statics of P between two equilibria, say c and $c^* > c$, will be determined by whether p/p' increases or decreases between c and c^*. As a result, if, for instance, p/p' is first increasing and then decreasing over the domain of optimal care levels, then the optimal number of redundancies will be at an intermediate level, as in Example 1 below.

Psychologists have long noted that people working in groups tend to expend less effort than people working as individuals, with larger groups exhibiting more "social loafing."[24] This finding corresponds to i) above. They have also observed that the introduction of automatic devices leads to a decrease in human performance, which corresponds to i) above.[25] Skitka et al. (2000) put subjects in simulated cockpits with imperfect automated monitoring aids. They then compared the performance of one-person crews with the performance of two-person crews. Although one might naively expect two-person crews to be much more likely to detect system irregularities than one-person crews, they found essentially no difference in detection rates, which is consistent with ii) (albeit in a relatively neutral way).

The following examples illustrate some interesting features of Theorem 4. In the first example, the optimal number of care-takers is an intermediate value.

Example 1 $S = [0,1]$, $p_a D = 40, p(c) = 1 - \frac{5}{4}c + \frac{1}{2}c^2$. *For any k, the symmetric equilibrium c_k solves $-p_a D \left(1 - \frac{5}{4}c_k + \frac{1}{2}c_k^2\right)^{k-1} \left(c_k - \frac{5}{4}\right) = 1$. The accident minimizing number of people is given by*

$$\arg\min_k P(k, q, p, p_a) = 5$$

In the second example, technological considerations restrict p_a to the interval $\left[\frac{1}{2}, 1\right]$. The probability of an accident is minimized by choosing the least reliable automation within this set.

[24] Pschologists' explanations for social loafing include arousal reduction, decreased evaluation potential, and a matching of anticipated decreased effort on the part of others (see Karau and Williams (1993) for a review).

[25] Psychologists' explanations include automation bias, and automation induced complacency. Consistent with ii), Skitka et. al (1999) find that experimental subjects are less reliable at detecting errors when aided by an automatic system. On the other hand, Parasuraman et al. (1993) conduct an experiment in which they find that the variability in the reliability of an automated system, but not the absolute value of this reliability, affect performance, a finding which is not consitent with ii) (although the interpretation of this finding is confounded by the fact that subjects were not given the reliability parameters).

Example 2 $S = [0,1], D > 2, p(c) = (1-c)^b, \ 1 \leq b < \frac{k+1}{k}, \ p_a \in \left[\frac{1}{2}, 1\right]$. *For any* p_a, *the symmetric equilibrium is* $c = 1 - (bDp_a)^{\frac{1}{1-bk}}$.

$$\underset{p_a \in \left[\frac{1}{2},1\right]}{\arg\min} P(k, q, p, p_a) = 1$$

The third example shows that our model is formally a generalization of the Volunteer's Dilemma (Samuelson (1984) and Diekmann (1985)). In this dilemma, an event can be prevented if and only if at least one of k people takes a costly action.

Example 3 *Each individual's payoff is given by:*

	Someone Else Acts	*No One Else Acts*
Takes Action	-1	-1
No Action	0	$-D$

In the symmetric mixed strategy equilibrium of this game, the probability of an event is monotonically increasing in k. *This result can be viewed as a special case of Example 2. To see this, set* $b = 1, p_a = 1$. *Then, a mixed strategy* $(\alpha, 1-\alpha)$ *in the Volunteer's Dilemma corresponds to a pure strategy* $c = \alpha$ *in Example 2. Since the equilibrium is interior, and* $\frac{p}{p'} = c - 1$ *is an increasing function, (ii) in Theorem 4 yields the Dilemma result that* P *is increasing in* k. *Since Darley and Latané (1968) introduced the concept of "diffusion of responsibility" into the psychology literature, this type of prediction has been tested often, with varying results (see Goeree, Holt and Moore (2005) and the references therein).*

Our results are also reminiscent of the "voluntary provision of public goods" literature. It has long been known that the provision of public goods is subject to a free-rider problem, and since Olson (1965) it has been argued that the severity of the problem increases with the number of individuals in society. Several authors have produced examples where the ratio between the optimal amount of a public good and the equilibrium amount of a voluntary provision game increases with the number of players. Gaube (2001) gives general sufficient conditions for this effect.[26] As in Gaube, we give sufficient conditions for the problem of

[26] Cornes (1993) analyzes the case in which the public good is produced via a Constant Elasticity of Substitution production function in which inputs are individual contributions. This case covers the standard case, plus other interesting cases. He does not analyze the effect of increasing the number of individuals.

underprovision to be exacerbated as n increases, but in addition we give sufficient conditions for the converse result to hold: we provide sufficient conditions under which the amount of the public good provided is increasing in n. In several other respects, our model is not comparable to this literature. In particular, in voluntary provision models, the public good is generally assumed to be the sum of the contributions c_i, whereas in our model it is $(1 - p_a \pi_{i=1}^{n} p(c_i))$, and we consider what happens to the aboslute level of the public good, not just the ratio to the optimal amount.

For a redundancy system requiring supervision, we may expect that, on the one hand, even a supervisor putting in a minimal amount of effort might detect an anomaly, while on the other hand, even a supervisor putting in a maximal amount might miss an anomaly. Formally, for $S = [0, M]$, this translates to $0 < p(M) < p(0) < 1$. Since $p(0) < 1$, the accident-minimizing number of people is then infinity. In practice, however, the "optimal" number of people will be less than infinity, for both technological reasons and economic reasons. As this section emphasizes, there may well be a non-monotonic relationship between the number of people and the probability of an accident, so that the optimal number of people is not necessarily the "constrained largest." At the same time, since $p(M) > 0$, the optimal number of people is unlikely to be one, in contrast with the Volunteer's Dilemma.

2 Checklists

Recent work in the health care industry suggests that the use of simple checklists may significantly reduce morbidity, mortality, and medical errors. For instance, Haynes et al. (2009) finds that implementing a surgical safety checklist in eight hospitals reduced the death rate from 1.5% to 0.8% and in-patient complications from 11% to 7% (see also Gawande (2010)).[27] However, despite the apparent success of checklists, many doctors resist their implementation.

It is not completely understood by what mechanisms checklists operate, or what constitutes the key elements of a checklist. Gawande (2010) suggests several possible benefits of checklists, including that they serve as simple reminders not to forget important steps, that they propose a better procedure than the one previously in place, that they foster a conversation amongst personnel, and that they encourage people to speak up about potential

[27] Checklists are used in many industries, and have a long history in aviation.

problems. With regard to the last suggestion, the "career concerns" model in the working version of this paper, Benoît and Dubra (2007), suggests one reason such an encouragement may be necessary. In a nutshell, suppose that an agent in a subordinate position, such as a nurse in surgery or a co-pilot on a plane, observes a possible problem that no one else has noticed. Say the agent believes the probability there is a problem to be $\frac{1}{100}$. For a critical problem, this probability is high enough to warrant reporting. At the same time, however, the chances are overwhelming that the agent's concerns will prove unfounded. If the agent is worried about appearing to be incompetent, and suffering attendant consequences, he, or she, has an incentive not to report his concerns; there will be underreporting of unlikely, but critical, possible problems. Many checklist protocols alleviate this problem by taking a "time out" in which all agents are expressly encouraged to air any concerns.

In this paper we focus on a particular aspect of checklists, namely the redundancies found in many of them. Amongst other things, the checklist used in Haynes et al. (2009) calls for the patient, surgeon, anesthesia professional, and nurse all to confirm the patient's identity. Kwaan et al. (2006) studies 16 surgical site-verification protocols, and finds that the number of redundant checks ranges from 5 to 20, averaging 12. A *redundancy checklist* calls for k different agents to carry out essentially the same check, in the hope that at least one of the checks is successful. This is captured by our model, with $p(c_i)$ being the probability that step i is successfully carried out.[28]

The adoption of a checklist – for instance, a series of oral verifications of the correct site for surgery – does not in and of itself guarantee that much attention is being paid during a particular step. Indeed, a well-known criticism of checklists is that steps may be carelessly addressed or even skipped altogether; Theorem 4 points in this direction. Nonetheless, a checklist can be expected to increase vigilance, as it reduces the (marginal) cost of care, since a certain degree of attention must be paid when a step is called out, and skipping a step takes some effort. Moreover, a checklist may induce greater care by instilling feelings of discipline and team involvement. In a very direct way, then, a checklist may reduce the probability of an accident and may even result in the accident probability that a non-equilibrium analysis, which neglects strategic slackening, anticipates. But is this reduction in probability a good thing?

[28]We note, however, that our model assumes independance of probabilities, which will not always hold in practice.

Thus far, the description of our model has not included a discussion of the nature of the agents' objective functions, that is, the agents' perception of the damage done by an accident and the source of the cost of preventative care. On the one hand, agents may be completely "altruistic" and internalize all aspects of the problem. For instance, suppose the agents have multiple job tasks and the cost of care is the attention taken away from attending to another task. Moreover, the agents and the principal agree that the damage from an accident is D.[29] Then we can write the principal's objective function as $-\bar{\theta}' p_a \pi_{i=1}^{k} p(c_i) D - \sum_{i=1}^{k} c_i$, where $\bar{\theta}'$ is the principal's estimate of the probability of an event. If the agents and the principal agree on the probability of an event (that is $\bar{\theta}' = \bar{\theta}_q$), then, it is easy to see that the symmetric equilibrium level of care is also the symmetric optimum from both the principal and the agents' point of view.[30] Then, a redundancy checklist results in too much care according to both the principal and the agents. However, if the agents are subject to the good news is bad effect[31] and have a lower estimate of the probability of an event, the agents' underinvest relative to the principal's optimum. In this case, a well-designed checklist leads to an improvement in care levels from the principal's point of view, but leads to wasteful care from the agents' point of view.

On the other hand, the agents' perception of the damage may be lower than the principal's and the cost of care may reflect things that the principal values less than the agents, such as attention taken away from personal tasks. With either of the possibilities, even if everyone agrees on the probability of an event, the agents will invest less than the principal deems optimal. Then a principal may find the checklist to be desirable while the agents find it undesirable.

Although Theorems 1 and 4 focus on symmetric equilibria, there may be asymmetric equilibria as well. Consider the game with the following parameters (analyzed in the appendix): $S = [0,1], D = 90, p(c) = (1 - .99c), \theta \equiv p_a \equiv 1$. This game can be interpreted as a volunteer's dilemma (see Example 3), in which a person who is attempting to take an

[29]Under a different interpretation than altruistic agents, the principal has chosen a payment function with large rewards for agents when no disaster occurs. Macdonald and Marx (2001) give (unrelated) reasons why a principal might choose such a payment function.

[30]That is, the optimum level of care, given the number of steps/agents. Whether or not this number was optimally chosen is another matter.

[31]Simon (2007) notes that orthopedic surgeons who have not had accidents are the least likely to implement a protocol designed to minimize wrong-site surgery.

action may nonetheless fail one out of one hundred times. As we vary the number of players, in the symmetric equilibrium of the game the probability of an accident starts at 1% with one player, hits a minimum of slightly above 0.01% with two players, and rises thereafter, approaching 1.12% in the limit. There are also asymmetric equilibria. All of these involve some of the players choosing zero care, and the rest of them choosing the same level of care. With respect to all equilibria, the probability of an accident is minimized when there are two or more players, two of the players choose the same level of care as in the two-player symmetric equilibrium, and all other players choose zero care.[32] The equilibrium which maximizes the sum of the agents' utilities, involves only one agent taking care. Once the presence of asymmetric equilibria is recognized, a new danger arises. If agents attempt to play to different asymmetric equilibria, they may end up in an out of equilibrium situation in which too many agents are taking no care, or one in which too much care is taken. A checklist may focus the agents on a particular equilibrium and, in this way, be desirable from both the principal's and the agents' point of view.

3 Conclusion

The world is a risky place, but how risky is a matter of some choice. Safeguards and backups can be built into nuclear power plants, planes can be extensively tested and regularly inspected, chemical facilities can have overlapping safety checks. Yet, though an ounce of prevention may be worth a pound of cure, that ounce is often missing. Inadequate care can be the result of miscalculations and misguided objectives. Thus, many analyses of the Challenger disaster have emphasized the increasing pressure to launch brought about by the commercialization of the Space Shuttle. We have shown, however, that these lapses can also be the result of a rational calculus by altruistic agents, leading to *imprevention* rather than prevention.

[32] Thus, when considering all equilibria, increasing the number of players beyond two does not increase the minimum probability of an accident (even though $\frac{p}{p'}$ is strictly increasing), but it is wasteful if there is any opportunity cost to these players.

4 Appendix

Proof of Theorem 1. We first show the existence of at least one symmetric Markov equilibrium. Since players' strategies depend only on the beliefs about the probability of an event, and effort does not affect these beliefs, in any given period player j maximizes that period's payoff, $-\overline{\theta}_{\mu} p_a \pi_{i=1}^{k} p\left(c_i\right) D - c_j$, where μ are the beliefs and c_i is player i's effort.

If S is bounded, let $c^* = \max S$. If $-\overline{\theta}_{\mu} p_a p^{k-1}\left(c^*\right) p'\left(c^*\right) D - 1 \geq 0$ then $c_i = c^*$ for all i is a symmetric equilibrium.

We now simultaneously analyze the case

$$-\overline{\theta}_{\mu} p_a p^{k-1}\left(c^*\right) p'\left(c^*\right) D - 1 < 0, \tag{1}$$

and the case of unbounded S. For the latter case, pick c^* sufficiently large that (1) is satisfied (such a c^* exists, because p bounded below implies that p' converges to 0).

If $-\overline{\theta}_{\mu} p_a p^{k-1}\left(0\right) p'\left(0\right) D - 1 \leq 0$, then $c_i = 0$ for all i is an equilibrium. If on the contrary, $-\overline{\theta}_{\mu} p_a p^{k-1}\left(0\right) p'\left(0\right) D - 1 > 0$, by equation (1), there is a $\widetilde{c} \in \left(0, c^*\right)$ such that $-\overline{\theta}_{\mu} p_a p^{k-1}\left(\widetilde{c}\right) p'\left(\widetilde{c}\right) D - 1 = 0$, and $c_i = \widetilde{c}$ for all i is an equilibrium.

We now show that there is exactly one symmetric equilibrium. Suppose that $c = 0$ is a symmetric equilibrium. Then $-\overline{\theta}_{\mu} p_a p^{k-1}\left(0\right) p'\left(0\right) D - 1 \leq 0$. Since for any other $c > 0$, we have $-\overline{\theta}_{\mu} p_a p^{k-1}\left(c\right) p'\left(c\right) D - 1 < -\overline{\theta}_{\mu} p_a p^{k-1}\left(0\right) p'\left(0\right) D - 1 \leq 0$, and $c_i = c$ is not a symmetric equilibrium.

If $c = \underline{c}$ is an interior symmetric equilibrium, then

$$-\overline{\theta}_{\mu} p_a p\left(\underline{c}\right)^{k-1} p'\left(\underline{c}\right) D = 1$$

must hold. Then $p' < 0 \leq p''$ ensures that there is no other c for which $-\overline{\theta}_{q} p_a p\left(c\right)^{k-1} p'\left(c\right) D = 1$, so there is no other symmetric equilibrium.

If $c = c^*$ is an equilibrium for the case of bounded S, $-\overline{\theta}_{\mu} p_a p^{k-1}\left(c^*\right) p'\left(c^*\right) D \geq 1$ holds, and for all $c < c^*$, we obtain $-\overline{\theta}_{\mu} p_a p^{k-1}\left(c\right) p'\left(c\right) D > 1$ so there is no other symmetric equilibrium. ■

Proof of Theorem 2. Claim 1. If two densities q' and q are such that q'/q is strictly increasing on their support $[0,1]$, then, for all $x \in (0,1)$, their cumulative distribution functions are such that $Q'\left(x\right) < Q\left(x\right)$. To see this, let $\overline{x}$ be such that $q'\left(\overline{x}\right) = q\left(\overline{x}\right)$. Then, for all $x \in \left(0, \overline{x}\right)$ we have $q'\left(x\right) < q\left(x\right)$ and so $Q'\left(x\right) < Q\left(x\right)$. For $x > \overline{x}$, $Q'\left(x\right) - Q\left(x\right)$

is increasing in x, since the derivative is strictly positive, and therefore $Q'(x) - Q(x) < Q'(1) - Q(1) = 0$.

Claim 2. $Q_n(\theta \leq x) > Q(\theta \leq x)$ for all $x \in (0,1)$. Note that by Bayes' Rule, the density of the posterior Q_n is

$$q_n(\theta) = \frac{\Pr(n \mid \theta)\Pr(\theta)}{\Pr(n)} = \frac{(1-\theta)\,q(\theta)}{\int_0^1 (1-z)\,q(z)\,dz}$$

so that the likelihood ratio of q and q_n is

$$\frac{q(\theta)}{q_n(\theta)} = \frac{\int_0^1 (1-z)\,q(z)\,dz}{1-\theta}$$

which is strictly increasing in θ. By Claim 1, $Q_n(\theta \leq x) > Q(\theta \leq x)$.

Thus, q strictly first order stochastically dominates q_n and $\int_0^1 \theta q_n(\theta)\,d\theta < \int_0^1 \theta q(\theta)\,d\theta$.

Claim 3. If $\overline{\theta}_r < \overline{\theta}_q$, $c(k,q,p,p_a) \geq c(k,r,p,p_a)$ with strict inequality if $c(k,q,p,p_a)$ is interior. Note that if $c(k,q,p,p_a) := c$ is interior, then $-\overline{\theta}_q p_a p(c)^{k-1} p'(c)\,D = 1$. For $\overline{\theta}_r < \overline{\theta}_q$ we have, $-\overline{\theta}_r p_a p(c)^{k-1} p'(c)\,D < 1$, and $-\overline{\theta}_r p_a p(c)^{k-1} p'(c)\,D$ decreasing in c implies $c(k,q,p,p_a) > c(k,r,p,p_a)$. If $c(k,q,p,p_a) = 0$, then for all $c' > 0$ we have $-\overline{\theta}_q p_a p^{k-1}(c')\,p'(c')\,D - 1 < 0$ which implies $-\overline{\theta}_r p_a p^{k-1}(c')\,p'(c')\,D - 1 < 0$, so that $c(k,r,p,p_a) = 0$ is the unique equilibrium. Finally, if $c(k,q,p,p_a) = c^*$ is an equilibrium for $S = [0,c^*]$, necessarily $c(k,q,p,p_a) \geq c(k,r,p,p_a)$. ■

Proof of Theorem 3. We first show that for all $t > T^*$, we have $\overline{\theta}_{q_{n^t}} < \overline{\theta}_{\widetilde{q}_{n^t}}$. Notice that after t draws of n, the posteriors of the agent and the principal are $B(a+t,b)$ and $B(d+t,e)$, respectively. We have,

$$\overline{\theta}_{q_{n^t}} < \overline{\theta}_{\widetilde{q}_{n^t}} \Leftrightarrow \frac{b}{a+b+t} < \frac{e}{d+e+t} \Leftrightarrow t > \frac{bd-ae}{e-b},$$

as needed. The claim about $c(k,q_{n^t},p,p_a) < c(k,\widetilde{q}_{n^t},p,p_a)$ follows by using $r = q_{n^t}$ and $q = \widetilde{q}_{n^t}$ in Theorem 2. ■

Proof of Theorem 4. Proof of i). Suppose that $k' > k$. If $c(k,q,p,p_a) := c$ is interior, then $-\overline{\theta}_q p_a p(c)^{k-1} p'(c)\,D = 1$. Therefore, $-\overline{\theta}_q p_a p(c)^{k'-1} p'(c)\,D < 1$, and since $-\overline{\theta}_q p_a p(\cdot)^{k'-1} p'(\cdot)\,D$ is decreasing, we obtain $c(k',q,p,p_a) < c(k,q,p,p_a)$. If $c(k,q,p,p_a) = 0$, then $-\overline{\theta}_q p_a p(0)^{k-1} p'(0)\,D \leq 1$ and $-\overline{\theta}_q p_a p(0)^{k'-1} p'(0)\,D < 1$, so that $c(k',q,p,p_a) = 0$. If $c(k,q,p,p_a) = c^*$ is an equilibrium for $S = [0,c^*]$, then necessarily $c(k',q,p,p_a) \leq c(k,q,p,p_a)$.

Suppose that $p_a' > p_a$. If $c(k,q,p,p_a) := c$ is interior, then $-\overline{\theta}_q p_a p(c)^{k-1} p'(c) D = 1$. We have $-\overline{\theta}_q p_a' p(c)^{k-1} p'(c) D > 1$, and $-\overline{\theta}_q p_a' p(c)^{k-1} p'(c) D$ decreasing in c implies $c(k,q,p,p_a') > c(k,q,p,p_a)$. If $c(k,q,p,p_a) = 0$, then necessarily $c(k,q,p,p_a') \geq c(k,q,p,p_a)$. If $c(k,q,p,p_a) = c^*$ is the equilibrium for $S = [0,c^*]$, then $-\overline{\theta}_q p_a p(c^*)^{k-1} p'(c^*) D \geq 1$, so that $-\overline{\theta}_q p_a' p(c^*)^{k-1} p'(c^*) D > 1$ and $c(k,q,p,p_a') = c^*$ is the equilibrium.

Proof of ii). If there is $\overline{c} \in S$ such that $-p(0)^{k-1} p'(0) > \frac{1}{\overline{\theta}_q p_a D} > -p(\overline{c})^{k-1} p'(\overline{c})$, there exists c such that $-p(c)^{k-1} p'(c) = \frac{1}{\overline{\theta}_q p_a D}$ holds; then $c(k,q,p,p_a) = c$ is the unique symmetric equilibrium (by Theorem 1).

Fix any $k' > k$ and let $c(k',q,p,p_a) := c'$. We now show that $P(k',q,p,p_a) > P(k,q,p,p_a)$ whenever p/p' is strictly increasing. From the proof of i), $c' < c$. Since c is interior, the first order condition implies $P(k,q,p,p_a) = -\frac{p(c)}{p'(c)D}$. Then, $p(\cdot)/p'(\cdot)$ strictly increasing implies

$$P(k,q,p,p_a) = -\frac{p(c)}{Dp'(c)} < -\frac{p(c')}{Dp'(c')} \leq P(k',q,p,p_a).$$

The proof for $\frac{p(c')}{p'(c')} > \frac{p(c)}{p'(c)}$ follows similarly. ■

The following Theorem relates to the game $S = [0,1]$, $D = 90$, $p(c) = 1 - .99c$ and $\theta \equiv p_a \equiv 1$ discussed in Section 2.

Theorem 5 *Suppose $(c_1, ..., c_k)$ is an equilibrium and let $I = \{i : c_i > 0\}$. Then $|I| \geq 1$. If $|I| = 1$ then $c_i = 1$ for $i \in I$. If $|I| > 1$, then*

$$c_i = \left[1 - \left(\frac{10}{9 \times 99}\right)^{\frac{1}{|I|-1}}\right] \frac{100}{99}$$

for all $i \in I$. The equilibrium probability of accident is $P = \left(\frac{10}{9\times 99}\right)^{\frac{|I|}{|I|-1}}$.

Proof. Given $(c_1, ..., c_k)$, player i's payoff is $u(c_i) = -90\left(1 - \frac{99}{100}c_i\right)\prod_{j\neq i}\left(1 - \frac{99}{100}c_j\right) - c_i$ and $u'(c_i) = 90 \times \frac{99}{100}\prod_{j\neq i}\left(1 - \frac{99}{100}c_j\right) - 1$. If $c_j = 0$ for all $j \neq i$, then $u'(c_i) > 0$, so that $|I| \geq 1$.

If some player j chooses $c_j = 1$, then for $i \neq j$, $u'(c_i) < 0$, so that all $i \neq j$ choose $c_i = 0$. From above, if all $i \neq j$ choose $c_i = 0$, then i chooses $c_j = 1$. Thus, if $|I| = 1$ then $c_i = 1$ for $i \in I$.

Suppose that $|I| > 1$. From the previous paragraph, $c_i < 1$ for all $i \in I$. Since $0 < c_i < 1$, we have $u'(c_i) = 0$. Hence $\prod_{j\neq i} p(c_j) = \frac{10}{9\times 99}$ for all $i \in I$. This implies that, for all $h, i \in I$, $p(c_h) = p(c_i) = \left(\frac{10}{9\times 99}\right)^{\frac{1}{|I|-1}}$, and $c_h = c_i = \left[1 - \left(\frac{10}{9\times 99}\right)^{\frac{1}{|I|-1}}\right]\frac{100}{99}$. ■

References

[1] Bénabou, R. (2008), "Groupthink: Collective Delusions in Organizations and Markets," mimeo, Princeton University.

[2] Benoît, J-P and J. Dubra (2007), "The Problem of Prevention," SSRN abstract id 906783.

[3] Benoît, J-P and J. Dubra (2011), "Apparent Overconfidence," *Econometrica* **79(5)**, 1591–1625.

[4] Born, P. and K. Viscusi (2006), "The catastrophic effects of natural disasters on insurance markets," *Journal of Risk and Uncertainty*, **33(1-2)**, 55-72.

[5] Burdett, K. and T. Vishwanath (1988): "Declining Reservation Wages," *Review of Economic Studies*, **55**, 655-65.

[6] Cornes, R, (1993) "Dyke Maintenance and Other Stories: Some neglected Types of Public Goods," *Quarterly Journal of Economics,* **108**, 259-71.

[7] Darley, J. and B. Latané (1968), "Bystander Intervention in Emergencies: diffusion of responsibility," *Journal of Personality and Social Psychology*, **8**, 377-83.

[8] Davies, D.R., and Parasuraman, R., (1981) , "The Psychology of Vigilance," Academic Press, London.

[9] Downer, J. (2011*a*) , "When Failure is an Option: Redundancy, reliability and regulation in complex technical systems," *Accounting Organizations and Society* **36(4-5)**, pp 269-83.

[10] Downer, J. (2011b), ""737-Cabriolet": The Limits of Knowledge and the Sociology of Inevitable Failure," *American Journal of Sociology* **117(3)**, pp 725-62.

[11] Downer, J. (2010), "On audits and airplanes: Redundancy and reliability-assessment in high technologies," LSE, CARR Discussion paper 61.

[12] Diekmann, A., (1985), "Volunteer's Dilemma," *Journal of Conflict Resolution*, **29** (4), 605-10.

[13] Dubra, J., (2004) “Optimism and Overconfidence in Search,” *Review of Economic Dynamics*, Volume **73**, Issue 1.

[14] Feynman, R., (1988), “What Do You Care What Other People Think?” W.W. Norton and Company, New York.

[15] Gaube, Thomas, (2001), “Group size and free riding when private and public goods are gross substitutes,” *Economics Letters* **70**, 127-132.

[16] Gawande, A. (2010), “The Checklist Manifesto,” Profile Books LTD, London.

[17] Goeree, J. C. Holt and A. Moore, (2005), “An Experimental Examination of the Volunteer’s Dilemma,” mimeo.

[18] Gollier, Christian (2002), “Optimal Prevention of Unknown Risks: a Dynamic Approach with Learning,” mimeo University of Toulouse.

[19] Haynes, A.B., Thomas G. Weiser, William R. Berry, Stuart R. Lipsitz, Abdel-Hadi S. Breizat, Patchen Dellinger, Teodoro Herbosa, Sudhir Joseph, Pascience L. Kibatala, Marie Carmela M. Lapitan, Alan F. Merry, Krishna Moorthy, Richard K. Reznick, Bryce Taylor, and Atul A. Gawande (2009), “A Surgical Safety Checklist to Reduce Morbidity and Mortality in a Global Population,” *The New England Journal of Medicine*, **360**, pp. 491-9.

[20] Heal, Geoffrey and H. Kunreuther, (2007), “Modeling Interdependent Risks,” *Risk Analysis*, **27(3)**, 612-34.

[21] Institute of Medicine (2000) “To Err is Human, Building a Better Safer Health Care System,” National Academy Press, Washington, D.C.

[22] Karau, S., and Williams, K. (1993), “Social Loafing: A Meta-Analytic Review and Theoretical Integration,” *Journal of Personality and Social Psychology*, **65**, 681-706.

[23] Kemeny, John, Bruce Babbitt, Patrick E. Haggerty, Carolyn Lewis, Paul A. Marks, Cora B. Marrett, Lloyd McBride, Harry C. McPherson, Russell W. Peterson, Thomas H. Pigford, Theodore B. Taylor, Anne D. Trunk (1979), “Report of the President’s Commission on the Accident at Three Mile Island.” Washington, D.C., Government Printing Office.

[24] Kunreuther, H. and G. Heal (2003), “Interdependent Security,” *The Journal of Risk and Uncertainty*, **26:2/3**; 231–49.

[25] Kwaan M.R., D.M. Studdert, M.J. Zinner and A.A. Gawande (2006) , “Incidence, patterns, and prevention of wrong-site surgery,” *Archives of Surgery,* **141(4)**, pp. 353-7.

[26] Lee, P.M. (1989), *Bayesian Statistics: An Introduction*, Ch. 3. New York: Oxford University Press.

[27] MacDonald, G. and L. Marx (2001), “Adverse Specialization,” *Journal of Political Economy*, **109(4)**, 864-99.

[28] Olson, Mankur, (1965), *The Logic of Collective Action*. Cambridge/Mass.: Harvard University Press.

[29] Parasuraman, R., Molloy, R., and Singh, L. (1993), “Performance Consequences of Automation-Induced “Complacency,”” *International Journal of Aviation Psychology*, 3(1), 1-23.

[30] Perrow, C. (1999), “Normal Accidents,” Princeton University Press, Princeton, New Jersey.

[31] Reason, J., (1990), Human Error, Cambridge University Press, Cambridge, United Kingdom

[32] Rothschild, M. (1974), “Searching for the Lowest Price When the Distribution of Prices is Unknown,” *Journal of Political Economy*, **82**, No. 4.

[33] Sagan, S. (2004), “The Problem of Redundancy Problem: Why More Nuclear Security Forces May Produce Less Nuclear Security,” *Risk Analysis*, **24(4)**.

[34] Samuelson, W., (1984), Potential Entry and Welfare, Mimeo.

[35] Simon, J.W. MD (2007), “Preventing Surgical Confusions in Ophtalmology (An American Ophtalmological Society Thesis), *Trans Am Ophthalmol Soc*, **105**, 513–529.

[36] Skitka, L., Mosier, K., and Burdick, M., (1999) - “Does Automation Bias Decision-Making?” *International Journal of Human-Computers Studies*, **51**, 991-1006.

[37] Skitka, L., Mosier, K., and Burdick, M., and Rosenblatt, B., (2000) "Automation Bias and Errors: Are Crews Better Than Individuals?" *International Journal of Aviation Psychology*, **10(1)**, 85-97.

[38] Vaughan, D. (1996), "The Challenger Launch Decision", The University of Chicago Press, Chicago.

[39] Viscusi, K. (1984), "The Lulling Effect: The Impact of Child-Resistant Packaging on Aspirin and Analgesic Ingestions," *The American Economic Review*, **74(2)**, 324-7.