

## ON THE PROBLEM OF PREVENTION\*

BY JEAN-PIERRE BENOÎT AND JUAN DUBRA<sup>1</sup>

*London Business School, U.K.;*  
*Universidad de Montevideo, Uruguay*

Disasters are often precipitated by insufficient preventive care. We argue that there is a problem of prevention in that this lack of care often stems from agents' rational calculations. Positive experiences lead agents to underestimate the risks of disasters; technological improvements and redundancies designed for safety induce agents to reduce their care. Although lower care increases the chances of an accident, the number of redundancies can be adjusted to offset this. However, the accident probability remains constant even as ostensible improvements in safety are made. Checklists can be used to decrease the number of accidents.

A remarkable number of disasters and near-disasters, from the nuclear mishap at Three Mile Island,<sup>2</sup> to the Union Carbide plant tragedy in Bhopal,<sup>3</sup> to the Challenger disaster,<sup>4</sup> to Hurricane Katrina<sup>5</sup> 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 postevent reporting. In a similar vein, a study by the Institute of Medicine (2000) concluded that each year over 44,000 people die in U.S. 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 article, 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 agents to underestimate these risks and underinvest in preventative care.
- (2) Redundancies designed for safety may induce agents to lessen the care they take.

As a consequence of these reductions in care, a system may become less safe even as it appears to be getting safer. *If* these effects are properly understood, this diminution in safety can be offset by an appropriate choice of the number of redundancies. However, the existence of these countervailing forces means that some attempts at safety improvement will have no impact.

\*Manuscript received June 2008; revised December 2011.

<sup>1</sup> We thank Martín Besfamille, Federico Echenique, Emeric Henry, Ennio Staccetti, and Federico Weinschelbaum for their comments. Benoît acknowledges the support of the C.V. Starr Center at NYU, and Dubra thanks the support from ANII through the Fondo Clemente Estable. Please address correspondence to: Juan Dubra, Department of Economics, Universidad de Montevideo, Prudencio de Pena 2440, Montevideo, 11.600, Uruguay. Phone: 598 98 005826. E-mail: [dubraj@um.edu.uy](mailto:dubraj@um.edu.uy).

<sup>2</sup> In March 1979, there was a partial meltdown of the reactor core of the Three Mile Island Unit 2 nuclear power plant.

<sup>3</sup> In December 1984, methyl isocyanate gas was released at the Union Carbide chemical plant in Bhopal, India, resulting in thousands of deaths and hundreds of thousands of injuries.

<sup>4</sup> The American space shuttle Challenger exploded shortly after takeoff on January 28, 1986.

<sup>5</sup> Hurricane Katrina struck southeast Louisiana on August 29th. Considerable damage was caused, including the flooding of 80% of New Orleans. There were over 1,500 deaths as a result.

When the potential losses from an accident are large, agents take more care, making high-loss systems less prone to accidents. The net effect on expected losses is ambiguous. If the number of redundancies is adjusted optimally, expected damages remain constant.

We establish the above results using a model of accident prevention that captures the aforementioned two factors. The model also incorporates the more familiar problems that occur when agents fail to fully internalize the costs of an accident. On the other hand, the model ignores other elements that may be germane. In particular, agents may have extraneous concerns, be subject to group-reinforced biases (as in Bénabou, 2008), or simply make mistakes (Reason, 1990, studies various types of errors to which humans are prone). Although these are important ingredients in explaining many accidents, we focus our attention on difficulties that remain even when actors are well motivated and well trained to avoid mistakes.

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 (2011b) argues that there is a category of *epistemic accidents*, which result from flawed theories and judgments. Sagan (2004) and Downer (2011a) highlight some of the same issues that we discuss, among other issues, but argue informally. We will return to this literature and to the relevant economics literature at various points in the article.

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) and we apply our analysis to checklists.

1. THE MODEL

To fix our ideas, consider a machine with one critical part, which may become defective and fail in any period with some given unknown probability. In each period, before 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 model.

There are  $k \geq 1$  agents, an automated device, and nature. In each period  $t = 0, 1, 2 \dots$ , 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  $\bar{\theta}_q = \int \theta dq(\theta)$ .

In every period, each agent chooses an investment in care  $c \in S = [0, M]$ , for some  $M \in \mathbf{R}_{++}$ . The upper bound  $M$  could, for instance, represent the agent concentrating fully on the task at hand. The choice of care is private information. If an agent invests  $c$  in care, with probability  $p(c)$  he or she fails to detect (and fix) an event that has occurred. The function  $p$  is twice continuously differentiable with  $p' < 0$ ,  $p'' \geq 0$ , and, when  $k = 1$ , the more restrictive  $p'' > 0$ . The additional assumption  $p'' > 0$  guarantees uniqueness of the equilibrium when  $k = 1$  (see Theorem 1).

There is also an automated device that 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 (although allowing the agents to continue would not change our results).

Given a profile of effort choices  $\mathbf{c} = (c_1, \dots, c_k)$  in the current period, and an expected probability of an event  $\bar{\theta}_q$ , the (subjective) probability of an accident, that is, the probability of

an undetected event, is  $\bar{\theta}_q p_a \pi_{i=1}^k p(c_i)$ .<sup>6</sup> An accident causes a loss of  $D$  to an agent; the payoff in any single period in which there is no accident is normalized to zero. Thus, the expected payoff of agent  $j = 1, \dots, k$  in the current period is  $-\bar{\theta}_q p_a \pi_{i=1}^k p(c_i)D - c_j$ .

We focus on Markov strategies. Specifically, we define the state to be agents' beliefs about  $\hat{\theta}$  and consider strategies that 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  $G(k, q, p_a, D)$ . We focus on symmetric equilibria, though we briefly discuss asymmetric equilibria in Section 3.

**THEOREM 1.** *The game  $G(k, q, p_a, D)$  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.

## 2. GOOD NEWS CAN BE BAD

Consider agents' beliefs about the inherent safety of their environment. That is, consider their beliefs about  $\hat{\theta}$ , the probability of an event. Scientific and other considerations yield a priori estimates that must be continually updated in the light of experience. Some industries, such as the airplane 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 consensus on the ways in which these need to be tested.<sup>7</sup> Other enterprises, such as nuclear power plants and the space shuttle, involve relatively new technologies with limited experience. These sparse 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 24 successful shuttle launches without a failure, and estimates of a catastrophic failure ranged from 1 in 100 to 1 in 100,000 (Feynman, 1988).<sup>8</sup> Similarly, before 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 (especially so-called blockbuster disasters; see Born and Viscusi, 2006).

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)

<sup>6</sup> We assume that the probabilities that different parts of the system fail are independent of each other. Downer (2011a) contends that in some cases it is difficult to tell whether or not independence holds.

<sup>7</sup> Nonetheless, Downer (2011b) argues that some innovations in airplane design, such as the introduction of new composite material, may be poorly understood.

<sup>8</sup> It should be noted, however, that the (management) estimate of 1 in 100,000 is a little hard to rationalize. Bénabou (2008) argues that the estimate is a result of "groupthink."

With regard to the Challenger, as part of the investigating commission, Feynman (1988) 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. (p. 223)

The slow shift toward a decreasing safety factor can be seen in many [areas]. (p. 230)

Vaughan (1996) has termed this steady decline in standards the “normalization of deviance,” though she ascribes a different mechanism to this decline than we do. This reduction in care has similarities to what Sagan (2004) terms agents *overcompensating* for increases in safety by taking additional risks.

The following theorem formalizes this phenomenon. Given a prior  $q$  about  $\theta$ , let  $q_n$  denote the Bayesian posterior beliefs following a period in which no event has occurred.<sup>9</sup> Let  $c(k, \bar{\theta}_q, p_a, D)$  denote the individual level of care in the symmetric equilibrium of  $G(k, q, p_a, D)$ .

**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,  $\bar{\theta}_{q_n} \equiv \int_0^1 \theta q_n(\theta) d\theta < \int_0^1 \theta q(\theta) d\theta = \bar{\theta}_q$ . The level of care taken is also smaller. That is,  $c(k, \bar{\theta}_{q_n}, p_a, D) \leq c(k, \bar{\theta}_q, p_a, D)$ , with strict inequality if  $c(k, \bar{\theta}_q, p_a, D)$  is interior.*

Thus, a string of periods with no events leads to both a reduced belief in the probability of an event and a decline in care. The net effect of these two changes on the *subjective* probability of an accident depends on  $p/p'$ , as detailed in Theorem 4 of the next section.

Although the 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, whereas (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 a divergence in the principal and the agent’s motivations.<sup>10</sup> Here, we focus on a different problem—one that arises from a discrepancy in the beliefs of the principal and the agent. We call this type of problem a *belief-based agency problem*.<sup>11</sup>

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. Although 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.

<sup>9</sup> Recall that the agent observes an event even if the event does not turn into an accident. Under an alternate formulation, the agent only observes events that turn into accidents. Adopting this alternate formulation would add an inferential complication.

<sup>10</sup> See, for instance, MacDonald and Marx (2001) for a principal–agent approach to accident prevention.

<sup>11</sup> 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 behavior by informing citizens of the risks involved, but typically finds that individuals’ beliefs concerning these risks can only be influenced, not dictated.

Formally, suppose the principal has belief  $\tilde{q}$ , whereas the agents have belief  $q$ . Both  $\tilde{q}$  and  $q$  are assumed to be represented by Beta distributions.<sup>12</sup> 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 nonevents, the agents are always more optimistic than the principal. 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 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.<sup>13</sup> If the principal and agents begin with the same mean belief, then the agents will begin underinvesting following the first nonevent.

Let  $q_{n^t}$  be the (Bayesian) posterior of  $q$  following  $t$  observations of  $n$ , and recall that  $\bar{\theta}_{q_{n^t}}$  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,  $\tilde{q}$ , are distributed according to  $B(d, e)$ . If the beliefs of the agent are more disperse than those of the principal, then enough nonevents 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  $\bar{\theta}_{q_{n^t}} < \bar{\theta}_{\tilde{q}_{n^t}}$ . For all  $t > T^*$ ,  $c(k, \bar{\theta}_{q_{n^t}}, p_a, D) \leq c(k, \bar{\theta}_{\tilde{q}_{n^t}}, p_a, D)$  and if  $c(k, \bar{\theta}_{\tilde{q}_{n^t}}, p_a, D)$  is interior, the inequality is strict.*

In the extreme, the principal's priors are so tight that she 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 adverse incident is very large, the optimal number of accidents is close to zero. For this reason, nuclear reactors are built so that a string of nonevents is the norm. Unfortunately, our results indicate that this success is, to some extent, self-defeating. The Kemeny Commission reaches much the same conclusion about "overupdating" on the part of power plant operators, writing in its report:

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. (p. 9; emphasis added)

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

<sup>12</sup> 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 nonevent it becomes  $B(\alpha + 2, \beta + 1)$ .

<sup>13</sup> 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.

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. At least two factors exacerbate this difficulty. The first one is the presence 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 particular plants. The second factor is the so-called availability heuristic (see Tversky and Kahneman, 1973). It has been argued that when estimating probabilities, people tend to place undue weight on factors that they can readily recall, chief among these being their personal experience.<sup>14</sup>

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  $\tilde{q}$  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 caretaking interact in nontrivial 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 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.

### 3. REDUNDANCIES

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.<sup>15</sup> 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.<sup>16</sup>

Although the meandering mind of a lifeguard may prove lethal, the danger posed pales in comparison with the potential harm from a nuclear or chemical plant. For this reason, these plants are designed so that the failings of a single individual are not sufficient for a disaster to

<sup>14</sup> 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 article.

<sup>15</sup> 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.

<sup>16</sup> See Davies and Parasuraman (1981) for a survey.

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)

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.<sup>17</sup> 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.<sup>18</sup> 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.<sup>19</sup>

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. However, why did the operators fail to check the gauges?<sup>20</sup> 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 that enhanced the safety of the plant also reduced the incentive of agents to take care, thus limiting the degree of safety that could be achieved.<sup>21</sup> An estimation of the safety of the system that neglects this strategic slackening will badly miss the mark.

Although strategic reductions in care raise the probability of a disaster and 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 overall effect of adding redundancy features is an increase in the probability of a disaster. Theorem 4 summarizes these findings.

Recall that  $c(k, \bar{\theta}_q, p_a, D)$  is the individual level of care in the symmetric equilibrium of  $G(k, q, p_a, D)$ , and let  $P(k, \bar{\theta}_q, p_a, D)$  be the equilibrium probability of an accident.

<sup>17</sup> 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 modeling, Perrow is concerned with dynamic difficulties. Nonetheless, he concedes that the number of failures that must take place for an accident to occur in and of itself provides a crucial measure of safety.

<sup>18</sup> 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.

<sup>19</sup> A somewhat related 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.

<sup>20</sup> Similar lapses in care have been noted at numerous other accident sites, including Three Mile Island.

<sup>21</sup> Sagan (2004) and Downer (2011a) argue informally that redundancies may lead to decreasing care.

THEOREM 4. *In the unique symmetric equilibrium of  $G(k, q, p_a, D)$ ,*

- (i)  *$c$  is decreasing in  $k$  and increasing in  $p_a, \bar{\theta}_q$ , and  $D$ —strictly if  $c$  is interior.*
- (ii)  *$P$  is decreasing in  $D$  and may be increasing or decreasing in its other arguments.*  
*Consider  $k' > k, D' > D$  and  $\bar{\theta}_q p'_a > \bar{\theta}_q p_a$ , and suppose the equilibrium is interior (i.e.,  $-p(0)^{k-1} p'(0) > \frac{1}{\bar{\theta}_q p_a D} > -p(M)^{k-1} p'(M)$ ).*  
*If  $\frac{p}{p'}$  is strictly increasing, then  $P(k', \bar{\theta}_q, p_a, D) > P(k, \bar{\theta}_q, p_a, D) > P(k, \bar{\theta}_q, p'_a, D)$  and  $P(k, \bar{\theta}_q, p_a, D)D > P(k, \bar{\theta}_q, p_a, D')D'$ ;*  
*if  $\frac{p}{p'}$  is strictly decreasing, then  $P(k', \bar{\theta}_q, p_a, D) < P(k, \bar{\theta}_q, p_a, D) < P(k, \bar{\theta}_q, p'_a, D)$  and  $P(k, \bar{\theta}_q, p_a, D)D < P(k, \bar{\theta}_q, p_a, D')D'$ .*

An example for which  $p/p'$  is increasing is  $p(c) = (1 - ac)^b$ , with  $a, b > 0$ ; an example for which it is decreasing is  $p(c) = a(1 + c)^{-\gamma}$ , for  $a, \gamma > 0$ .

Theorem 4 tells us that a system that is inherently unsafe may have fewer accidents than a relatively safe system.<sup>22</sup> 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, suppose  $S = \{0, 1\}$ ,  $p(0) = 1$ ,  $p(1) = 0$ , and  $p_a = 1$ . For small enough  $\bar{\theta}_q > 0$ , it is optimal for the agent to take no care, resulting in an accident probability of  $\bar{\theta}_q$ , whereas for large enough  $\bar{\theta}_q$  it is optimal for the agent to take perfect care, resulting in an accident probability of 0.

From Theorem 4, agents will take greater care when damages are greater so that the probability of an accident will be smaller in high-loss systems. The net effect on expected damages is ambiguous, however. Note that the damage  $D$  in the theorem is, in fact, the agents’ perception of the damage. If this is smaller than society’s perception, the agents will invest less in care than the social optimum, even if this optimum fully incorporates the agents’ costs. The agents’ perception of the damage might be low because the agents ignore the externalities of an accident, among other reasons.

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 equilibrium care levels, then the accident-minimizing 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.”<sup>23</sup> This finding corresponds to point (i) above. They have also observed that the introduction of automatic devices leads to a decrease in human performance, which corresponds to point (i) above.<sup>24</sup> 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 point (ii) (albeit in a relatively neutral way).

<sup>22</sup> Viscusi (1984) argues that child safety caps on aspirin led to a decrease in adult care. Although he offers no theoretical argument on the net safety impact, his empirical analysis suggests that decreases in care offset the benefit of the safety cap.

<sup>23</sup> Psychologists’ 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).

<sup>24</sup> Psychologists’ explanations include automation bias and automation induced complacency. Consistent with point (ii), Skitka et al. (1999) find that experimental subjects are less likely to detect 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, affects performance, a finding that is not consistent with point (ii) (although the interpretation of this finding is confounded by the fact that subjects were not given the reliability parameters).

The following examples illustrate some interesting features of Theorem 4. In the first example, the optimal number of caretakers assumes an intermediate value.

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

$$\arg \min_k P(k, \bar{\theta}_q, p_a, D) = 5.$$

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

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

$$\arg \min_{p_a \in [\frac{1}{2}, 1]} P(k, \bar{\theta}_q, p_a, D) = 1.$$

The third example shows that our model is formally a generalization of the Volunteer’s Dilemma (Samuelson, 1984; Diekmann, 1985). In this dilemma, an accident 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  $b = 1$ ,  $\bar{\theta}_q p_a = 1$ . Then, a mixed strategy  $(\alpha, 1 - \alpha)$  in the Volunteer’s Dilemma corresponds to a pure strategy  $c = \alpha$  in Example 2. Because 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 et al., 2005, and the references therein).

Although we focus on symmetric equilibria throughout this article, there may be asymmetric equilibria as well. In the following example, which is analyzed in the appendix, we consider all equilibria.

EXAMPLE 4.  $S = [0, 1]$ ,  $\bar{\theta}_q p_a D = 90$ ,  $p(c) = 1 - .99c$ . This game can be interpreted as a Volunteer’s Dilemma in which a person who acts still fails to prevent an accident 1% of the time. As the number of players increases, 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 positive 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. Thus, when considering all equilibria, increasing the number of players beyond two does not *strictly* increase the minimum probability of an accident, even though  $\frac{p}{p'}$  is strictly increasing. However, increasing

the number of players beyond two does not reduce the probability of an accident either, and it is wasteful if there is any opportunity cost to these players. Note also that 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.

Our results are 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.<sup>25</sup> As in Gaube, we give sufficient conditions for the problem of underprovision to be exacerbated as  $k$  increases, but in addition we give sufficient conditions for the converse result to hold; that is, we provide sufficient conditions under which the amount of the public good provided is increasing in  $k$ . In several other respects, our model is not comparable with 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  $1 - p_a \pi_{i=1}^k p(c_i)$ , and we consider what happens to the absolute level of the public good, not just the ratio to the optimal amount.

#### 4. ACCIDENT MINIMIZATION

In this section, we analyze the maximal safety of the system.

Consider a principal who, by making investments, can exert *some* control over the probability of an event,  $\bar{\theta}_q$ , and the quality of the automated device,  $p_a$ . At the same time, she is free to choose the number of agents/redundancies. Suppose the cost of the investments and the cost of the agents are negligible, so that the principal seeks to minimize the probability of an accident. That is, for any  $\bar{\theta}_q p_a$  that can be feasibly attained, she chooses  $k$  to minimize  $P(k, \bar{\theta}_q, p_a, D)$ .

Under some conditions, the probability of an accident can be taken to 0 by hiring an infinite number of agents. (One such condition is that agents can detect an accident even when taking zero care ( $p(0) < 1$ )). However, when the number of agents goes to infinity, the assumptions of the model become strained, as does the notion that the cost of hiring agents is negligible. Although the theorems in this section are formally correct under these conditions, they are more meaningful when these conditions do not hold and the optimal number of agents is finite.

We can think of  $\bar{\theta}_q p_a$  as measuring (agents' beliefs about) the technological safety of the system. If  $\bar{\theta}_q p_a \approx 1$ , the system relies almost entirely on agents to prevent an accident. If there is a single agent, he will take maximal care if the loss from an accident is important enough—more precisely, if  $D > -\frac{1}{p'(M)}$ . As  $\bar{\theta}_q p_a$  is lowered, the technological safety of the system improves. If the system became safe enough, even a lone agent would find it optimal to take less than maximal care, and, for an extremely safe system would go so far as to take zero care. However, for the type of systems we are primarily interested in, the technological limitations are such that and the damage is so large that a single agent would always take maximal care (formally,  $\bar{\theta}_q p_a D > -\frac{1}{p'(M)}$ ). Put differently, we are mainly interested in situations in which, if agents exert less than maximal effort, it is because of the redundancies that have been introduced to compensate for the agents' inherent fallibilities.<sup>26</sup>

For such situations, Theorem 5 below shows that when the number of agents is adjusted optimally, increases in the technological safety of the system are exactly offset by reductions in care. That is, the minimum accident probability remains constant as  $\bar{\theta}_q p_a$  decreases, so that these improvements do nothing for the safety of the system. These improvements may nonetheless

<sup>25</sup> 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.

<sup>26</sup> We discuss what happens when a single agent would not take maximal care following Theorem 5 below.

be valuable, as they enable a reduction in the number of agents (see Theorem 6 below) and a concomitant cost savings. At the same time, Theorem 5 shows that the diminution in safety that results from agents overupdating following good news (Theorem 3) can, in principle, be offset by an appropriate change in the number of agents.

Although changes in the technological safety do not affect the maximal safety of the system, changes in the damages, or the agents' perception of the damages, do. Theorem 5 establishes that systems with greater potential damages,  $D$ , have a lower minimum accident probability. However, minimum expected damages are constant.

Recall that the equilibrium probability of an accident is given by  $P(k, \bar{\theta}_q, p_a, D) = \bar{\theta}_q p_a p(c(k, \bar{\theta}_q, p_a, D))^k$ , where  $c(k, \bar{\theta}_q, p_a, D)$  is the equilibrium probability of care. From the proof of Theorem 1 in the appendix, we have that

$$(1) \quad c(k, \bar{\theta}_q, p_a, D) = \begin{cases} 0 & \text{if } -\bar{\theta}_q p_a p(0)^{k-1} p'(0) D \leq 1 \\ M & \text{if } -\bar{\theta}_q p_a p(M)^{k-1} p'(M) D \geq 1 \\ c : \bar{\theta}_q p_a p(c)^{k-1} = -\frac{1}{p'(c)D} & \text{otherwise} \end{cases}$$

We now take the number of agents  $k$  to be any positive real number, not just an integer, and use (1) to define  $c(k, \bar{\theta}_q, p_a, D)$ . Restricting  $k$  to integer values, as we have done up to now, only results in a more complicated statement of Theorem 5, with bounds on probability differences.

**THEOREM 5.** *Suppose that  $D' < D$ . For all  $-\frac{1}{p'(M)D} < \bar{\theta}_q p_a, \bar{\theta}_q' p_a'$ , the minimum probability of an accident is constant in  $\bar{\theta}_q$  and  $p_a$  and decreasing in  $D$ . That is,  $\inf_k P(k, \bar{\theta}_q, p_a, D) = \inf_k P(k, \bar{\theta}_q', p_a', D)$  and  $\inf_k P(k, \bar{\theta}_q, p_a, D) \leq \inf_k P(k, \bar{\theta}_q, p_a, D')$ . The minimum expected loss is constant in  $D$ . That is,  $\inf_k P(k, \bar{\theta}_q, p_a, D)D = \inf_k P(k, \bar{\theta}_q, p_a, D')D'$ .*

Theorem 5 follows from the fact that, when the accident minimizing care that agents expend is interior, the minimum expected loss is given by  $\min_{c \in (0, M)} -\frac{p}{p'}$ . As an illustration of the theorem, consider Example 1 with  $D = 50$ . For all  $-\frac{1}{p'(M)D} = 0.08 < \bar{\theta}_q p_a$ ,  $\inf_k P(k, \bar{\theta}_q, p_a, D) = 0.012$ .

Moving beyond the parameters of the theorem, for  $0.03 \leq \bar{\theta}_q p_a \leq 0.08$ , we still have  $\inf_k P(k, \bar{\theta}_q, p_a, D) = 0.012$ . Thus, for 97% of the possible parameter range, the minimal accident probability remains constant. As  $\bar{\theta}_q p_a$  falls from 0.03 to 0.016,  $\inf_k P(k, \bar{\theta}_q, p_a, D)$  starts to increase until it reaches 0.016 at  $\bar{\theta}_q p_a = 0.016$ . Finally, as  $\bar{\theta}_q p_a$  falls further,  $\inf_k P(k, \bar{\theta}_q, p_a, D)$  also falls and, in fact, equals,  $\bar{\theta}_q p_a$ . In this final range, agents take no care in equilibrium. This pattern is quite general. In particular, decreases in  $\bar{\theta}_q p_a$  reduce the minimal accident probability below the constant value the minimum probability takes when  $\bar{\theta}_q p_a > -\frac{1}{p'(M)D}$ , only if these decreases are drastic enough to induce agents to take no care.

Staying with Example 1, for  $\bar{\theta}_q p_a = \frac{4}{5}$ , as  $D$  increases from just above 5 to infinity, the minimum probability of an accident falls continually from 0.12 to 0 in the limit. The expected damage, however, remains constant at 0.6.

We now consider changes in the accident-minimizing number of agents. Theorem 2 indicates that following periods of nonevents, agents will reduce their care, whereas Theorem 4 implies that as the agents' perception of the loss caused by an accident falls, agents again take less care. From Theorem 5, the principal can offset these reductions by an appropriate choice of redundancies. On the face of it, the principal could accomplish this offset either by increasing the number of agents, in order to compensate for the fall in care, or reducing the number of agents, in order to induce each agent to take more care. Theorem 6 below shows that, in the present situation, the principal does the latter.

Given  $(\bar{\theta}_q, p_a, D)$ , let  $k(\bar{\theta}_q, p_a, D)$  be the smallest number of agents that results in the minimal accident probability, or infinity when hiring ever more agents keeps reducing the probability

of an accident. Recall that  $\bar{\theta}_{q_{nt}}$  is the agents' estimated probability of an event following  $t$  observations of no event.

**THEOREM 6.** *If  $D' < D$  and  $-\frac{1}{p'(M)D'} < \bar{\theta}_{q_{nt}} p_a < 1$ , then  $k(\bar{\theta}_{q_{nt}}, p_a, D') \leq k(\bar{\theta}_q, p_a, D)$ .*

Returning to Example 1, with  $D = 50$  and  $p_a = 1$ , as  $\bar{\theta}_{q_{nt}}$  falls from just below 1 to 0.08, the accident-minimizing number of agents falls from 5 to 2. Fixing  $\bar{\theta}_q p_a = \frac{4}{5}$ , as  $D$  falls from 500 to 50, the accident-minimizing number of agents falls from 8 to 5.

### 5. CHECKLISTS

Recent research 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) found that the implementation of a surgical safety checklist in eight hospitals reduced the death rate from 1.5% to 0.8% and reduced in-patient complications from 11% to 7%. Checklists are used in many other industries as well, most notably in aviation (see Gawande, 2010, for a discussion).

It is not completely understood by what mechanisms checklists operate or what constitutes the key elements of a successful 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, and that they encourage people to speak up about potential problems.<sup>27</sup> In this section, 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 detects a problem if there is one. This is captured by our model with  $1 - p(c_i)$  being the probability that check  $i$  is successful.

The adoption of a checklist—for instance, a series of oral verifications of the correct site for surgery—does not by 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, which is consistent with Theorem 4. Nonetheless, by explicitly attracting an agent's attention to these steps, a checklist reduces the marginal cost of care. In fact, taking little care itself may require effort in the face of someone calling out a check. Thus, a surgeon who finds it too time-consuming to personally confirm a patient's identity absent a checklist may find that there is little (extra) cost to doing so when a nurse calls out for confirmation as one step in a checklist. This reduction in marginal cost will lead to extra care, as described in Theorem 7 below.

Let  $G(x)$  denote the game in which a redundancy checklist has reduced an agents' cost of supplying care  $c$  from the amount  $c$  to the amount  $f(c, x)$ , where  $f$  is a family of functions parametrized by  $x \in [0, 1]$ , with the property that for all  $x' > x$  and all  $c \in (0, M]$ , (a)  $f_1(c; x) > 0$  and  $f_{11}(c; x) \geq 0$ , and (b)  $f_1(c; x') \leq f_1(c; x)$ . Property (a) says that the cost of effort increases in a convex way with effort, whereas (b) says that increases in  $x$  reduce the marginal cost of effort. In the game  $G(x)$ , agent  $j$  maximizes  $-\bar{\theta}_q p_a \pi_{i=1}^k p(c_i) D - f(c_j; x)$ .

<sup>27</sup> With regard to the last suggestion, the "career concerns" model in the working version of this article, Benoit and Dubra (2007), indicates one reason why encouraging people to speak out may be necessary. 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 that 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 he has an incentive to not report his concerns. As a result, 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.

**THEOREM 7.** *There is a unique symmetric equilibrium of  $G(x)$  in Markov strategies. The equilibrium effort  $c$  is increasing in  $x$  and the equilibrium probability of an accident  $P$  is decreasing in  $x$ .*

Thus, a checklist may be used to counter strategic slackening on the part of agents as well as any reduced effort induced by the good news is bad effect. Although the traditional view is that the downside of a long checklist is the time spent on the checks, Theorems 4, 5, and 6 show that even absent time considerations a long checklist may not be desirable. A judicious choice of the number of checks is needed to ensure that agents are not induced to behave carelessly.

A checklist may also serve a coordination function. As discussed in Example 4 of Section 3, if there are multiple asymmetric equilibria and agents miscoordinate and attempt to play to different ones, they may end up in an out of equilibrium situation in which very little care is taken. A checklist may circumvent this problem by focusing the agents on a particular equilibrium.

## 6. 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 lapses in care can also be the result of a rational calculus by altruistic agents leading to *imprevention* instead of prevention.

## APPENDIX

*Proof of Theorem 7. Existence.* Because 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,  $-\bar{\theta}_q p_a \pi_{i=1}^k p(c_i)D - f(c_j; x)$ .

If  $-\bar{\theta}_q p_a p(M)^{k-1} p'(M)D - f_1(M; x) \geq 0$ , then  $c_i = M$  for all  $i$  is a symmetric equilibrium.

Suppose that  $-\bar{\theta}_q p_a p(M)^{k-1} p'(M)D - f_1(M; x) < 0$ . If  $-\bar{\theta}_q p_a p(0)^{k-1} p'(0)D - f_1(0; x) \leq 0$ , then  $c_i = 0$  for all  $i$  is an equilibrium. If, on the contrary,  $-\bar{\theta}_q p_a p(0)^{k-1} p'(0)D - f_1(0; x) > 0$ , then there is a  $\tilde{c} \in (0, M]$  such that  $-\bar{\theta}_q p_a p(\tilde{c})^{k-1} p'(\tilde{c})D - f_1(\tilde{c}; x) = 0$ , and  $c_i = \tilde{c}$  for all  $i$  is an equilibrium.

**Uniqueness.** We now show that there is exactly one symmetric equilibrium.

Suppose that  $c = 0$  is a symmetric equilibrium. Then  $-\bar{\theta}_q p_a p(0)^{k-1} p'(0)D - f_1(0; x) \leq 0$ . For any  $c > 0$ , we have

$$-\bar{\theta}_q p_a p(c)^{k-1} p'(c)D - f_1(c; x) < -\bar{\theta}_q p_a p(0)^{k-1} p'(0)D - f_1(0; x) \leq 0,$$

and  $c_i = c$  is not a symmetric equilibrium.

Suppose that  $c = M$  is a symmetric equilibrium. Then  $-\bar{\theta}_q p_a p(M)^{k-1} p'(M)D \geq f_1(M; x)$ . For all  $c < M$ , we have  $-\bar{\theta}_q p_a p(c)^{k-1} p'(c)D > f_1(c; x)$ , so there is no other symmetric equilibrium.

Suppose that  $c = \underline{c}$  is an interior symmetric equilibrium. Then  $-\bar{\theta}_q p_a p(\underline{c})^{k-1} p'(\underline{c})D = f_1(\underline{c}; x)$ . Note that  $f'' \geq 0$ ,  $p' < 0$ , and  $p'' \geq 0$  or  $k = 1$  and  $p'' > 0$  ensures that there is no other  $c$  for which  $-\bar{\theta}_q p_a p(c)^{k-1} p'(c)D = f_1(c; x)$ , so there is no other symmetric equilibrium.

**Comparative statics.** Let  $x' > x$ . If  $c_x = M$  is the equilibrium of  $G(x)$ , then

$$0 \leq -\bar{\theta}_q p_a p(M)^{k-1} p'(M)D - f_1(M; x) \leq -\bar{\theta}_q p_a p(M)^{k-1} p'(M)D - f_1(M; x')$$

so that  $c_x$  is also an equilibrium of  $G(x')$ .

Suppose, by way of contradiction, that  $c_x < M$  is the equilibrium of  $G(x)$  and  $c_{x'} < c_x$  is the equilibrium of  $G(x')$ . Because  $k > 1$  and  $p'' \geq 0$  or  $k = 1$  and  $p'' > 0$ , we have that

$$\begin{aligned} 0 &\geq -\bar{\theta}_q p_a p(c_{x'})^{k-1} p'(c_{x'}) D - f_1(c_{x'}; x') > -\bar{\theta}_q p_a p(c_x)^{k-1} p'(c_x) D - f_1(c_x; x') \\ &\geq -\bar{\theta}_q p_a p(c_x)^{k-1} p'(c_x) D - f_1(c_x; x), \end{aligned}$$

which is a contradiction.

Because  $P(x) = \bar{\theta}_q p_a p(c_x)^k$  and  $c_x$  is increasing in  $x$ ,  $P(x)$  is decreasing. ■

*Proof of Theorem 1.* Theorem 1 follows by setting  $f(c, x) = c$  in the above proof. ■

*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  $Q'(x) < Q(x)$ . To see this, let  $\bar{x} = \sup\{x : q'(x) \leq q(x)\}$ . Then, for all  $x \in (0, \bar{x})$  we have  $q'(x) < q(x)$  and so  $Q'(x) < Q(x)$ . For  $x > \bar{x}$ ,  $Q'(x) - Q(x)$  is increasing in  $x$ , because 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  $x \in (0, 1)$ . From Bayes' Rule, the density of the posterior  $Q_n$  is

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

so that

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

which is strictly increasing in  $\theta$ . 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.*  $c(k, \bar{\theta}_{q_n}, p_a, D) \leq c(k, \bar{\theta}_q, p_a, D)$ . This follows from Theorem 4. ■

*Proof of Theorem 3.* We first show that for all  $t > T^*$ , we have  $\bar{\theta}_{q_{n^t}} < \bar{\theta}_{\tilde{q}_{n^t}}$ . 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,

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

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

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

Suppose that  $\underline{p}'_a > p_a$ . If  $c(k, \bar{\theta}_q, p_a, D) := c$  is interior, then  $-\bar{\theta}_q p_a p(c)^{k-1} p'(c) D = 1$ . We have  $-\bar{\theta}_q p'_a p(c)^{k-1} p'(c) D > 1$ , and  $-\bar{\theta}_q p'_a p(c)^{k-1} p'(c) D$  decreasing in  $c$  implies  $c(k, \bar{\theta}_q, p'_a, D) > c(k, \bar{\theta}_q, p_a, D)$ . If  $c(k, \bar{\theta}_q, p_a, D) = 0$ , then necessarily  $c(k, \bar{\theta}_q, p'_a, D) \geq c(k, \bar{\theta}_q, p_a, D)$ . If  $c(k, \bar{\theta}_q, p_a, D) = M$  is the equilibrium, then  $-\bar{\theta}_q p_a p(M)^{k-1} p'(M) D \geq 1$ , so that  $-\bar{\theta}_q p'_a p(M)^{k-1} p'(M) D > 1$  and  $c(k, \bar{\theta}_q, p'_a, D) = M$  is the equilibrium. Similar arguments apply to  $\theta_{q'} > \bar{\theta}_q$  and  $D' > D$ .

Proof of (ii). That  $P$  is decreasing in  $D$  follows from  $c(k, \bar{\theta}_q, p_a, D)$  increasing in  $D$ ,  $p(\cdot)$  decreasing and  $P(k, \bar{\theta}_q, p_a, D) = \bar{\theta}_q p_a p(c(k, \bar{\theta}_q, p_a, D))^k$ .

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

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

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

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

Finally, for  $\bar{\theta}_{q'} p'_a > \bar{\theta}_q p_a$  let  $c'' := c(k, \bar{\theta}_{q'}, p'_a, D)$ , so that by (i)  $c'' > c$ . Repeating the steps of the previous paragraph we obtain that  $P(k, \bar{\theta}_q, p_a, D) > P(k, \bar{\theta}_{q'}, p'_a, D)$  when  $p/p'$  is strictly increasing and the reverse when  $p/p'$  is decreasing. This concludes the proof. ■

We now prove Theorems 5 and 6. Recall that throughout the article we assume  $p'' \geq 0$ , except that when  $k = 1$  we impose  $p'' > 0$  to guarantee uniqueness of the equilibrium. As this assumption is ambiguous when  $k$  is endogenously chosen, for Theorems 5 and 6 we assume only  $p'' \geq 0$  for all  $k$ . The condition  $-\frac{1}{p'(M)D} < \bar{\theta}_q p_a$  in the statements of the theorems guarantees uniqueness because choosing care level  $M$  is then the unique optimal thing for a lone agent to do.

In order to prove these theorems, we first present a definition and series of lemmas. If  $p(M) > 0$  and  $-\frac{1}{p'(M)D} < \bar{\theta}_q p_a$ , define  $k_M$  by

$$(A.1) \quad -\bar{\theta}_q p_a p(M)^{k_M-1} p'(M) D = 1.$$

LEMMA 11. *The function  $c(\cdot, \bar{\theta}_q, p_a, D)$  defined in Equation (1) is continuous for every  $(\bar{\theta}_q, p_a, D)$ .*

PROOF. It is easy to check that  $c(k, \bar{\theta}_q, p_a, D) = \arg \min_{c \in [0, M]} (1 + \bar{\theta}_q p_a p(c)^{k-1} p'(c) D)^2$  and that it is indeed a function. By the theorem of the maximum,  $c$  is upper hemicontinuous, and because it is a function, it is continuous. ■

LEMMA 12. *If  $p(0) = 1, p(M) > 0$ , and  $-\frac{1}{p'(M)D} < \bar{\theta}_q p_a$ ,  $\inf_k P(k, \bar{\theta}_q, p_a, D) = -\sup_k \frac{p(c(k, \bar{\theta}_q, p_a, D))}{p'(c(k, \bar{\theta}_q, p_a, D))D}$ .*

PROOF. Note that  $-\frac{1}{p'(M)D} < \bar{\theta}_q p_a$  implies that for every  $k, c(k, \bar{\theta}_q, p_a, D) > 0$ . For all  $\hat{k} < k_M$

$$P(k_M, \bar{\theta}_q, p_a, D) = \bar{\theta}_q p_a p(M)^{k_M} < \bar{\theta}_q p_a p(M)^{\hat{k}} = P(\hat{k}, \bar{\theta}_q, p_a, D).$$

Hence,  $\inf_k P(k, \bar{\theta}_q, p_a, D) = \inf_{k \geq k_M} P(k, \bar{\theta}_q, p_a, D)$ . From Equation (A.1),  $P(k_M, \bar{\theta}_q, p_a, D) = -\frac{p(M)}{p'(M)D}$ . For  $k > k_M$ ,  $c(k, \bar{\theta}_q, p_a, D) \in (0, M)$ , so that  $P(k, \bar{\theta}_q, p_a, D) = -\frac{p(c(k, \bar{\theta}_q, p_a, D))}{p'(c(k, \bar{\theta}_q, p_a, D))D}$ . Therefore,  $\inf_k P(k, \bar{\theta}_q, p_a, D) = -\sup_k \frac{p(c(k, \bar{\theta}_q, p_a, D))}{p'(c(k, \bar{\theta}_q, p_a, D))D}$ . ■

LEMMA 13. Suppose  $p(0) = 1$ ,  $p(M) > 0$ , and  $-\frac{1}{p'(M)D} < \bar{\theta}_q p_a$ . Then for every  $c \in (0, M]$  there is a unique  $k$  such that  $c(k, \bar{\theta}_q, p_a, D) = c$ , and  $P(c(k, \bar{\theta}_q, p_a, D)) = -\frac{p(c(k, \bar{\theta}_q, p_a, D))}{p'(c(k, \bar{\theta}_q, p_a, D))D}$ .

PROOF. For every  $c \in (0, M]$ , for large enough  $k$ ,

$$-p(c)^{k-1} p'(c) < \frac{1}{\bar{\theta}_q p_a D} < -p'(M) \leq -p'(c) = -p(c)^{1-1} p'(c).$$

Thus, for all  $c \in (0, M]$ , for some  $\bar{k}$ ,  $-p(c)^{\bar{k}-1} p'(c) = \frac{1}{\bar{\theta}_q p_a D}$  and  $c(\bar{k}, \bar{\theta}_q, p_a, D) = c$ .

Because  $p(c) > 0$ , for any  $k \neq \bar{k}$ ,  $-p(c)^{k-1} p'(c) \neq -p(c)^{\bar{k}-1} p'(c) = \frac{1}{\bar{\theta}_q p_a D}$ , establishing uniqueness. ■

LEMMA 14. If  $p(0) = 1$  and  $-\frac{1}{p'(M)D} < \bar{\theta}_q p_a$ , then, for all  $k$ ,  $P(k, \bar{\theta}_q, p_a, D) \geq -\frac{p(c(k, \bar{\theta}_q, p_a, D))}{p'(c(k, \bar{\theta}_q, p_a, D))D}$ .

PROOF. For all  $k$  such that  $c_k = c(k, \bar{\theta}_q, p_a, D) = 0$ ,  $P(k, \bar{\theta}_q, p_a, D) = \bar{\theta}_q p_a p(0)^k = \bar{\theta}_q p_a > -\frac{1}{p'(M)D} \geq -\frac{1}{p'(0)D} = -\frac{p(0)}{p'(0)D}$ .

For all  $k$  such that  $c_k \in (0, M)$ ,  $P(k, \bar{\theta}_q, p_a, D) = -\frac{p(c_k)}{p'(c_k)D}$ .

For all  $k$  such that  $c_k = M$ , we have that  $k \leq k_M$  and

$$P(k, \bar{\theta}_q, p_a, D) = \bar{\theta}_q p_a p(M)^k \geq \bar{\theta}_q p_a p(M)^{k_M} = -\frac{p(M)}{p'(M)D} = -\frac{p(c_k)}{p'(c_k)D}. \quad \blacksquare$$

*Proof of Theorem 5. Case 1.*  $p(M) = 0$ . For  $k = 1$ ,  $-\frac{1}{p'(M)D} < -\frac{1}{p'(M)D'} < \bar{\theta}_q p_a$  ensures  $c(1, \bar{\theta}_q, p_a, D) = M$  and  $P(k, \bar{\theta}_q, p_a, D) = 0$ , and we are done.

*Case 2.*  $p(M) > 0$ , and  $p(0) < 1$ . For large enough  $k$ ,  $c(k, \bar{\theta}_q, p_a, D) = 0$ . Hence,  $\lim_{k \rightarrow \infty} p(0)^k = 0$  so that  $\inf_k P(k, \bar{\theta}_q, p_a, D) = 0$ , and again we are done.

*Case 3.*  $p(M) > 0$ , and  $p(0) = 1$ . By Lemmas 12 and 13 we obtain

$$\inf_k P(k, \bar{\theta}_q, p_a, D) = -\sup_k \frac{p(c(k, \bar{\theta}_q, p_a, D))}{p'(c(k, \bar{\theta}_q, p_a, D))D} = -\sup_c \frac{p(c)}{p'(c)D},$$

which is constant in  $\bar{\theta}_q, p_a$  and decreasing in  $D$ . Similarly,  $\inf_k P(k, \bar{\theta}_q, p_a, D') = -\sup_c \frac{p(c)}{p'(c)D'}$ , which ensures  $\inf_k P(k, \bar{\theta}_q, p_a, D)D = -\sup_c \frac{p(c)}{p'(c)} = \inf_k P(k, \bar{\theta}_q, p_a, D')D'$ . ■

In Lemma 15 below we show that

$$(A.2) \quad k(\bar{\theta}_q, p_a, D) = \begin{cases} \infty & \text{if } p(0) < 1 \text{ or} \\ & 0 = \arg \max_c \frac{p(c)}{p'(c)} \\ \min\{k \in \arg \min_{k \geq 1} P(k, \bar{\theta}_q, p_a, D)\} & \text{otherwise} \end{cases}$$

is well defined and that  $k(\bar{\theta}_q, p_a, D)$  is a minimizer of  $P(k, \bar{\theta}_q, p_a, D)$ .

LEMMA 15. Assume  $-\frac{1}{p'(M)D} < \bar{\theta}_q p_a$ . If  $p(0) < 1$  or  $0 = \arg \max_c \frac{p(c)}{p'(c)}$ ,  $\inf_k P(k, \bar{\theta}_q, p_a, D) = \lim_{k \rightarrow \infty} P(k, \bar{\theta}_q, p_a, D)$ . Otherwise, there is a  $k^*$  such that  $\inf_k P(k, \bar{\theta}_q, p_a, D) = P(k^*, \bar{\theta}_q, p_a, D)$ .

PROOF. *Case 1.* If  $p(0) < 1$ , for all  $k$ ,  $p(c(k, \bar{\theta}_q, p_a, D)) < 1$ , and

$$0 \leq \inf_k P(k, \bar{\theta}_q, p_a, D) \leq \lim_{k \rightarrow \infty} \bar{\theta}_q p_a D p(c(k, \bar{\theta}_q, p_a, D))^k = \lim_{k \rightarrow \infty} P(k, \bar{\theta}_q, p_a, D) = 0.$$

*Case 2.*  $p(0) = 1$  and  $0 = \arg \max_c \frac{p(c)}{p'(c)}$ .

Because, for all  $k$ ,  $1 < -\bar{\theta}_q p_a p'(M)D < -\bar{\theta}_q p_a p'(0)D = -\bar{\theta}_q p_a p(0)^{k-1} p'(0)D$ , we have  $c(k, \bar{\theta}_q, p_a, D) > 0$ . Hence, for large enough  $k$ , we have  $0 < c(k, \bar{\theta}_q, p_a, D) < M$ , and  $\lim_{k \rightarrow \infty} P(k, \bar{\theta}_q, p_a, D) = \lim_{c \rightarrow 0} -\frac{p(c)}{p'(c)D}$ . Because  $0 = \arg \max_c \frac{p(c)}{p'(c)}$ , we have  $\lim_{c \rightarrow 0} -\frac{p(c)}{p'(c)D} \leq -\frac{p(c(k, \bar{\theta}_q, p_a, D))}{p'(c(k, \bar{\theta}_q, p_a, D))D}$  for all  $k$ . Hence, using Lemma 14,

$$\lim_{k \rightarrow \infty} P(k, \bar{\theta}_q, p_a, D) \leq -\frac{p(c(k, \bar{\theta}_q, p_a, D))}{p'(c(k, \bar{\theta}_q, p_a, D))D} \leq P(k, \bar{\theta}_q, p_a, D)$$

for all  $k$ , which proves that  $\lim_{k \rightarrow \infty} P(k, \bar{\theta}_q, p_a, D) = \inf_k P(k, \bar{\theta}_q, p_a, D)$ .

*Case 3.*  $p(0) = 1$  and  $\exists c^* > 0$ , such that  $c^* \in \arg \max_c \frac{p(c)}{p'(c)}$ .

If  $p(M) = 0$ , then, because  $c(1, \bar{\theta}_q, p_a, D) = M$ , we have  $P(1, \bar{\theta}_q, p_a, D) = 0$  and setting  $k^* = 1$  yields the desired result.

If  $p(M) > 0$ , by Lemma 13 there is a  $k^*$  such that  $c(k^*, \bar{\theta}_q, p_a, D) = c^*$ . Using Lemmas 12 and 13

$$\begin{aligned} \inf_k P(k, \bar{\theta}_q, p_a, D) &= -\sup_k \frac{p(c(k, \bar{\theta}_q, p_a, D))}{p'(c(k, \bar{\theta}_q, p_a, D))D} = -\sup_c \frac{p(c)}{p'(c)D} = -\frac{p(c^*)}{p'(c^*)D} \\ &= P(k^*, \bar{\theta}_q, p_a, D). \end{aligned}$$

Lemma 15 shows that  $k(\bar{\theta}_q, p_a, D)$  is well defined. If neither  $p(0) < 1$  or  $0 = \arg \max_c \frac{p(c)}{p'(c)}$ , there is a  $k^*$  that minimizes  $P(k, \bar{\theta}_q, p_a, D)$ , so that  $\arg \min_{k \geq 1} P(k, \bar{\theta}_q, p_a, D)$  is well defined, and because  $c(k, \bar{\theta}_q, p_a, D)$  is continuous in  $k$  (Lemma 11) so is  $P(k, \bar{\theta}_q, p_a, D)$ , which ensures that  $\arg \min_{k \geq 1} P(k, \bar{\theta}_q, p_a, D)$  is closed.

*Proof of Theorem 6.* If  $p(0) < 1$  or  $0 = \arg \max_c \frac{p(c)}{p'(c)}$ , then, from Lemma 15,  $k(\bar{\theta}_{q_{nt}}, p_a, D') = k(\bar{\theta}_q, p_a, D) = \infty$ .

Assume therefore that  $p(0) = 1$  and let  $\bar{c} = \max_c \{c \in \arg \max \frac{p(c)}{p'(c)}\} > 0$ . If  $p(M) = 0$ , then  $P(1, \bar{\theta}_q, p_a, D) = P(1, \bar{\theta}_{q_{nt}}, p_a, D') = 0$ , so  $k(\bar{\theta}_{q_{nt}}, p_a, D') = k(\bar{\theta}_q, p_a, D) = 1$ , and we are done, so assume also  $p(M) > 0$ .

By Lemmas 12 and 13, there is a unique  $\bar{k}$  such that  $\bar{c} = c(\bar{k}, \bar{\theta}_q, p_a, D)$  and

$$\inf_k P(k, \bar{\theta}_q, p_a, D) = -\sup_c \frac{p(c)}{p'(c)D} = -\frac{p(\bar{c})}{p'(\bar{c})D} = -\frac{p(c(\bar{k}, \bar{\theta}_q, p_a, D))}{p'(c(\bar{k}, \bar{\theta}_q, p_a, D))D} = P(\bar{k}, \bar{\theta}_q, p_a, D).$$

Similarly, there is a unique  $\bar{k}'$  such that  $\bar{c} = c(\bar{k}', \bar{\theta}_{q_{nt}}, p_a, D')$  and

$$\begin{aligned} \inf_k P(k, \bar{\theta}_{q_{nt}}, p_a, D') &= -\sup_c \frac{p(c)}{p'(c)D'} = -\frac{p(\bar{c})}{p'(\bar{c})D'} = -\frac{p(c(\bar{k}', \bar{\theta}_{q_{nt}}, p_a, D'))}{p'(c(\bar{k}', \bar{\theta}_{q_{nt}}, p_a, D'))D'} \\ &= P(\bar{k}', \bar{\theta}_{q_{nt}}, p_a, D'). \end{aligned} \tag{A.3}$$

If  $\bar{c} = M$ , for all  $k < \bar{k}$ ,

$$\inf_k P(k, \bar{\theta}_q, p_a, D) = P(\bar{k}, \bar{\theta}_q, p_a, D) = \bar{\theta}_q p_a P(M)^{\bar{k}} < \bar{\theta}_q p_a P(M)^k = P(k, \bar{\theta}_q, p_a, D),$$

which shows that no  $k < \bar{k}$  minimizes  $P(k, \bar{\theta}_q, p_a, D)$ , and hence  $\bar{k} = k(\bar{\theta}_q, p_a, D)$ .

If  $\bar{c} < M$ , for all  $k < \bar{k}$ , because  $\bar{c}$  is interior,  $c(k, \bar{\theta}_q, p_a, D) > c(\bar{k}, \bar{\theta}_q, p_a, D) = \bar{c}$ , and, because,  $\bar{c} = \max_c \{c \in \arg \max \frac{p(c)}{p'(c)}\}$ , we have  $\frac{p(c(k, \bar{\theta}_q, p_a, D))}{p'(c(k, \bar{\theta}_q, p_a, D))} < \frac{p(\bar{c})}{p'(\bar{c})}$ . Lemma 14 then implies

$$P(k, \bar{\theta}_q, p_a, D) \geq -\frac{p(c(k, \bar{\theta}_q, p_a, D))}{p'(c(k, \bar{\theta}_q, p_a, D))D} > -\frac{p(\bar{c})}{p'(\bar{c})D} = P(\bar{k}, \bar{\theta}_q, p_a, D),$$

which shows that no  $k < \bar{k}$  minimizes  $P(k, \bar{\theta}_q, p_a, D)$ , and hence  $\bar{k} = k(\bar{\theta}_q, p_a, D)$ .

From Theorem 4

$$c(k(\bar{\theta}_q, p_a, D), \bar{\theta}_{q_{n^t}}, p_a, D') \leq c(k(\bar{\theta}_q, p_a, D), \bar{\theta}_q, p_a, D) = \bar{c} = c(\bar{k}, \bar{\theta}_{q_{n^t}}, p_a, D').$$

Applying Theorem 4 again, we have  $\bar{k}' \leq k(\bar{\theta}_q, p_a, D)$ . Finally, Equation (4) shows that  $k'$  minimizes  $P(k, \bar{\theta}_{q_{n^t}}, p_a, D')$ , so it implies that

$$k(\bar{\theta}_{q_{n^t}}, p_a, D') = \min_k \{k \in \arg \min_{k \geq 1} P(k, \bar{\theta}_{q_{n^t}}, p_a, D')\} \leq \bar{k}' \leq k(\bar{\theta}_q, p_a, D). \quad \blacksquare$$

The following proposition relates to the game in Example 4, discussed in Section 3. In that game,  $S = [0, 1]$ ,  $\bar{\theta}_q p_a D = 90$ ,  $p(c) = 1 - .99c$ .

**PROPOSITION 16.** *Suppose  $(c_1, \dots, c_k)$  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  $(c_1, \dots, c_k)$ , player  $i$ 's payoff is  $u(c_i) = -90(1 - \frac{99}{100}c_i) \prod_{j \neq i} (1 - \frac{99}{100}c_j) - c_i$  and  $u'(c_i) = 90 \times \frac{99}{100} \prod_{j \neq i} (1 - \frac{99}{100}c_j) - 1$ . 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  $i \in I$ . Because  $0 < c_i < 1$ , 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

BÉNABOU, R., "Groupthink: Collective Delusions in Organizations and Markets," Mimeo, Princeton University, 2008.  
 BENOÎT, J.-P., AND J. DUBRA, "The Problem of Prevention," SSRN abstract id 906783, 2007.  
 ———, AND ———, "Apparent Overconfidence," *Econometrica* 79(5) (2011), 1591–625.  
 BORN, P., AND K. VISCUSI, "The Catastrophic Effects of Natural Disasters on Insurance Markets," *Journal of Risk and Uncertainty*, 33(1–2) (2006), 55–72.  
 BURDETT, K., AND T. VISHWANATH, "Declining Reservation Wages," *Review of Economic Studies* 55 (1988), 655–65.

- CORNES, R., "Dyke Maintenance and Other Stories: Some Neglected Types of Public Goods," *Quarterly Journal of Economics* 108 (1993), 259–71.
- DARLEY, J., AND B. LATANÉ, "Bystander Intervention in Emergencies: Diffusion of Responsibility," *Journal of Personality and Social Psychology* 8 (1968), 377–83.
- DAVIES, D. R., AND R. PARASURAMAN, *The Psychology of Vigilance* (London: Academic Press, 1981).
- DIEKMANN, A., "Volunteer's Dilemma," *Journal of Conflict Resolution* 29(4) (1985), 605–10.
- DOWNER, J., "When Failure Is an Option: Redundancy, Reliability and Regulation in Complex Technical Systems," *Accounting Organizations and Society* 36(4–5) (2011a), 269–83.
- , "'737-Cabriolet': The Limits of Knowledge and the Sociology of Inevitable Failure," *American Journal of Sociology* 117(3) (2011b), 725–62.
- DUBRA, J., "Optimism and Overconfidence in Search," *Review of Economic Dynamics* 73(1) (2004), 198–218.
- FEYNMAN, R., *What Do You Care What Other People Think?* (New York: W.W. Norton and Company, 1988).
- GAUBE, T., "Group Size and Free Riding When Private and Public Goods Are Gross Substitutes," *Economics Letters* 70 (2001), 127–32.
- GAWANDE, A., *The Checklist Manifesto* (London: Profile Books LTD, 2010).
- GOEREE, J., C. HOLT, AND A. MOORE, "An Experimental Examination of the Volunteer's Dilemma," Mimeo, CalTech, 2005.
- GOLLIER, C., "Optimal Prevention of Unknown Risks: A Dynamic Approach with Learning," Mimeo, University of Toulouse, 2002.
- HAYNES, A. B., T. G. WEISER, W. R. BERRY, S. R. LIPSITZ, A.-H. S. BREIZAT, P. DELLINGER, T. HERBOSA, S. JOSEPH, P. L. KIBATALA, M. CARMELA, M. LAPITAN, A. F. MERRY, K. MOORTHY, R. K. REZNICK, B. TAYLOR, AND A. A. GAWANDE, "A Surgical Safety Checklist to Reduce Morbidity and Mortality in a Global Population," *The New England Journal of Medicine* 360 (2009), 491–9.
- HEAL, G., AND H. KUNREUTHER, "Modeling Interdependent Risks," *Risk Analysis* 27(3) (2007), 612–34.
- INSTITUTE OF MEDICINE, *To Err Is Human: Building a Better Safer Health Care System* (Washington, DC: National Academy Press, 2000).
- KARAU, S., AND K. WILLIAMS, "Social Loafing: A Meta-Analytic Review and Theoretical Integration," *Journal of Personality and Social Psychology* 65 (1993), 681–706.
- KEMENY, J., B. BABBITT, P. E. HAGGERTY, C. LEWIS, P. A. MARKS, C. B. MARRETT, L. MCBRIDE, H. C. MCPHERSON, R. W. PETERSON, T. H. PIGFORD, T. B. TAYLOR, AND A. D. TRUNK, "Report of the President's Commission on the Accident at Three Mile Island." Washington, DC, Government Printing Office, 1979.
- KUNREUTHER, H., AND G. HEAL, "Interdependent Security," *The Journal of Risk and Uncertainty*, 26(2/3) (2003), 231–49.
- KWAAN M. R., D. M. STUDDERT, M. J. ZINNER, AND A. A. GAWANDE, "Incidence, Patterns, and Prevention of Wrong-Site Surgery," *Archives of Surgery* 141(4) (2006), 353–7.
- LEE, P. M., *Bayesian Statistics: An Introduction* (New York: Oxford University Press, 1989).
- MACDONALD, G., AND L. MARX, "Adverse Specialization," *Journal of Political Economy* 109(4) (2001), 864–99.
- OLSON, M., *The Logic of Collective Action* (Cambridge, MA: Harvard University Press, 1965).
- PARASURAMAN, R., R. MOLLOY, AND L. SINGH, "Performance Consequences of Automation-Induced 'Complacency,'" *International Journal of Aviation Psychology* 3(1) (1993), 1–23.
- PERRON, C., *Normal Accidents* (Princeton, NJ: Princeton University Press, 1999).
- REASON, J., *Human Error*, (Cambridge, UK: Cambridge University Press, 1990).
- ROTHSCHILD, M., "Searching for the Lowest Price when the Distribution of Prices Is Unknown," *Journal of Political Economy* 82(4) (1974), 689–711.
- SAGAN, S., "The Problem of Redundancy Problem: Why More Nuclear Security Forces May Produce Less Nuclear Security," *Risk Analysis* 24(4) (2004), 935–46.
- SAMUELSON, W., "Potential Entry and Welfare," Working Paper 12/84, Boston University, School of Management, 1984.
- SKITKA, L., K. MOSIER, AND M. BURDICK, "Does Automation Bias Decision-Making?" *International Journal of Human-Computers Studies* 51 (1999), 991–1006.
- , ———, ———, AND B. ROSENBLATT, "Automation Bias and Errors: Are Crews Better than Individuals?" *International Journal of Aviation Psychology* 10(1) (2000), 85–97.
- TVERSKY, A., AND D. KAHNEMAN, "Availability: A Heuristic for Judging Frequency and Probability." *Cognitive Psychology* 5(1) (1973), 207–33.
- VAUGHAN, D., *The Challenger Launch Decision* (Chicago: The University of Chicago Press, 1996).
- VISCUSI, K., "The Lulling Effect: The Impact of Child-Resistant Packaging on Aspirin and Analgesic Ingestions," *The American Economic Review* 74(2) (1984), 324–7.