OPTIMAL GOVERNMENT POLICY TOWARD RISK IMPOSED BY UNCERTAINTY CONCERNING FUTURE GOVERNMENT ACTION

Louis Kaplow

Discussion Paper No. 20

8/86

Program in Law and Economics
Harvard Law School
Cambridge, MA 02138

The Program in Law and Economics is supported by a grant from the John M. Olin Foundation.

OPTIMAL GOVERNMENT POLICY TOWARD RISK IMPOSED BY UNCERTAINTY CONCERNING FUTURE GOVERNMENT ACTION

Louis Kaplow

Abstract

This paper examines policy responses to the problem presented by uncertainty concerning future government action -- which is widespread in many areas of government activity, such as taxation, regulation, and the allocation of budgetary resources. It models the effects of government compensation for losses (or windfall taxation of gains) and other modes of transition relief, such as grandfathering and phase-ins, when potentially affected investors have the option of obtaining private insurance. The analysis examines ex ante effects on incentives concerning investment and insurance decisions, emphasizing the moral hazard that arises both from private insurance and government transition relief.

This investigation is part of a larger project concerning the economic analysis of government transitions. This paper in large part consists of the formal analysis supporting the argument in Discussion Paper No.7, "An Economic Analysis of Legal Transitions;" 99 *Harvard Law Review* (1986).

Table of Contents

1. INTRODUCTION

An important but infrequently examined source of risk in the economy arises from uncertainty concerning future government policy. Although substantial effort has been devoted to understanding how financial markets will behave in light of expectations concerning government macroeconomic policy, little attention has been given to the problem of analyzing the significance of risk that is imposed by the prospect of government reforms or how government policies might be adjusted to mitigate the adverse effects of government-created risk.

Simple examples of existing risks include the possibility that the government will enact or repeal some tax or subsidy on a given activity, take property for a public project, change regulation for or deregulate an industry, or substantially alter its supply or demand for some good or service. Most directly, such reforms could impose substantial gains or losses on holders of the affected assets, and more generally, the wealth of other actors will often be affected by adjustments in government policy. As a result, in advance of such reforms, private entities will be affected by the risk that such changes might be forthcoming. A natural question, therefore, is whether the government should offer insurance, compensation, or other relief to mitigate the impact of its actions.

Feldstein (1976b) has argued that those who suffer losses as a result of government reforms should in fact be compensated. His analysis, however, fails to consider two important questions concerning the appropriateness of government intervention into operations of the market: why it might be that the market is unable to respond optimally and what adverse incentive effects might result from government compensation. This investigation focuses on the ex ante efficiency effects, which involve questions of risk bearing and incentives, which have received minimal treatment in the literature.[1] Such a focus, rather than more traditional discussions of horizontal equity, is suggested by the argument in Kaplow (1985a).

Little attention has been given to the question of whether government-created risks ought to be treated any differently from general market risks. The basic thesis of this investigation is that the two issues have very much in common. Although the usual view is that the appropriate treatment of market risks is provided by the market itself, in some contexts the desirability of government action has received attention. Depending upon what one thinks of the relative capabilities of government decisionmaking and decisionmaking through a decentralized market, widely differing positions on government policy toward market risk, and other market phenomena more generally, are possible. For the purposes of this study, I will simply assume that market mechanisms are usually preferable for the commonly stated reasons, particularly those concerning the ability of the market to process information.[2] The general conclusion in the base case is thus against transition policies designed to mitigate the impact of government reform. One implication of the analysis is that if existing intuitions in favor of mitigation are to be defended, one must look either to more complicated combinations of market failures or to institutional considerations that go beyond conventional analysis of optimal policy determination.

----

[1] Blume and Rubinfeld (1984) and Blume, Rubinfeld, and Shapiro (1984) examine takings. The former, informal article emphasizes the risk-spreading properties of compensation, suggesting that they are favorable, and the latter, formal paper focuses on incentive effects, suggesting the undesirability of compensation. As discussed in Kaplow (1986), they do not reconcile this tension. Watson (1984) examines some of the incentive effects, but does not offer welfare comparisons that incorporate risk-spreading effects. His article focuses on compensation that is only applicable to the uninsured portion of losses, and his discussion suggests that he has in mind losses causes by nature (or perhaps by the market) rather than by the government.

[2] To the extent one rejected this traditional assumption, the government would probably rely far less on tax, subsidy, and other approaches that utilize the market to make the ultimate allocative decisions. Such a shift in perspective would change not merely the answer to the questions addressed here but in many cases the questions themselves.

1.1 Problem Definition

Before proceeding with the investigation, some elaboration on the nature of the problem is helpful. First, it is important to state the source of the transition problem precisely, to avoid possible confusion about the issues at stake. Transition problems arise not from *change* itself, but rather from the fact that future government policy is *uncertain*. As a simple illustration, if it had been announced long ago that, as of next January 1, a different tax rate would apply to sales of a given commodity, any motivation for relief would vanish. Because the change is fully anticipated, all decisions have been able to take it into account every bit as well as they could have accounted for the tax rate if had been certain that there was not going to be any change.[3] No questions of risk or fairness would arise. Concerns that have been voiced concerning transition are motivated by the fact that prospects for change are generally uncertain.[4]

An important related implication is that the relevant set of "changes" that raise transition concerns is not limited to actual reforms. So long as there is not complete compensation, any change in the *prospects* for future reform should themselves be considered "reforms" for the purposes of most analysis of transitions. The reason is that market values will change in response to such events in the same manner that they change in response to reforms themselves; the only difference is one of degree.

With or without this last factor, it is clear that transitions are ubiquitous. Given the recent rate of tax reform and change in regulation, this statement hardly comes as a surprise. Yet such instances are only a fraction of the reforms that occur. Changes in government priorities,

----

[3] In fact, one could say that no "change" is involved, as the new tax rate on January 1 is part of the originally anticipated path of future policy. From this perspective, continuing the old rate would constitute a "change." The point is that "change" is too vague a concept, and uncertainty is necessary to raise the issues addressed here.

[4] One obvious exception, not considered here, is that adjustment cost considerations might be relevant to the optimal rate of change. As discussed in subsection 3.1, such factors are best understood as relating to determination of the optimal underlying policy, rather than "transition policy," as later discussion of this distinction should make clear.

reflected in government demand for goods and services, or in private demand, as influenced by transfer payments, equally constitute reforms. From the viewpoint of a private investor, an event that makes a project less valuable because of a new tax is much the same as one that causes the same impact through an equivalent decrease in demand caused by different budget priorities.[5] One must also include countless regulations and court decisions. With this latter group, even interpretations of regulations or statutes that do not "make new law" are relevant because, so long as there was not complete certainty ex ante, a clarifying ruling may affect the probability investors attach to certain regimes being applicable to their actions. Even changes in foreign policy have important impacts on many investments.

The ubiquity of government risk parallels that of other sources of risk that investors must take into account. In addition to natural disasters, or lesser changes in weather patterns, there is uncertainty concerning technology, actions of competitors, and consumer preferences. As will be explored below, one typically can find analogous market risks for virtually any government risk one can imagine. In addition, risks of actions by foreign governments, whether they relate to trade policy or the bombing of ships, are generally viewed as among the risks that investors must face. It will be seen that the analysis of transition relief for the most part directly parallels analysis that could be made of the desirability of government relief for the wide variety of market risks.

1.2 Organization of Investigation

The analysis proceeds as follows. Section 2 presents a two-stage model of moral hazard for analyzing the effect of government mitigation of losses caused by policy change.[6] The model views transition relief from an ex ante

----

[5] There are differences -- for example, if the tax is a function of the output of a pollutant that can be controlled, or if the demand shift affects the profitability of other opportunities in the same or related sectors -- but there is the common element that both can be unexpected and, even taking into account adjustments, can result in gains and losses of the same magnitude.

[6] For ease of continuing revision, equations are numbered consecutively within subsections. To minimize potential confusion that might result,

perspective, making it much like compulsory insurance. Investors, taking government policy as given, make their own insurance arrangements and investment decisions. The basic intuition is that transition relief, by transferring costs and benefits of investors' actions to the government, distorts investment and insurance decisions. For example, compensation for losses tends to produce excessive aggregate risk spreading and overinvestment. The base case model is followed by discussion that considers further the connection between government and market risk. It is then argued that windfall gains, which receive much less attention than transition losses, should be viewed symmetrically. For example, the ex ante effect of taxing gains -- mirroring the case of compensation for losses -- will be underinvestment. This section closes by considering special cases where moral hazard is less problematic in the absence of government relief: the risk-neutral case, with its relationship to firms where ownership is widely held, and instances where the insurance company has sufficient information to design policies that preserve incentives.

Section 3 extends the base case model. First, it demonstrates that alternative transition mechanisms -- such as compensation, grandfathering, delayed implementation, and phase-ins -- are largely similar in terms of ex ante effects. Then, the desirability of retroactive application in some contexts is assessed. Also considered are the implications of uncertainty concerning transition relief itself. Finally, an illustrative example involving multiple sources of moral hazard is examined to explore how more complex market imperfections might create a stronger case for transition relief in some instances.

Section 4 evaluates the base case assumptions. Additional imperfections, such as transaction costs, probability misperceptions, and adverse selection are considered, as well as the possibility that the government has superior information or risk-bearing capacity. Next, the section evaluates two key

----

equation numbers after the base case are preceded by an asterisk or other distinguishing notation, and all references to equation numbers are within the subsection unless the contrary is explicitly indicated. Propositions are also numbered within subsections. Those not part of the base case have abbreviations preceding the numbers to indicate the topic to which they refer.

assumptions used throughout: that transition policy is consistent over time and that the underlying reforms which raise the transition issue are themselves optimal. Finally, brief consideration is given to the relationship between the discussion here and issues concerning confiscatory taxation and income distribution.

Section 5 briefly explores how the analysis applies in a number of familiar settings: tort liability; takings; changes in government demand, government contracting, and initial program design; and some areas of tax reform. The purpose is not to offer a definitive resolution of the issues in any of these contexts, but rather to illustrate the basic approach and suggest some of the directions in which it might lead. Concluding remarks are in section 6.

2. A TWO-STAGE MODEL OF MORAL HAZARD

This section presents and discusses the model that will be the basis for the analysis in the remainder of this investigation. The first subsection presents the major assumptions that frame the inquiry. The second contains the base case, which models government transition relief as a simple insurance policy that covers a predetermined portion of losses (or gains) that might result from government policy change. This formulation corresponds most directly to pure compensation; alternative mechanisms will be considered in subsection 3.1. A "compensation premium," equaling the expected value of government payments, is included so that the government's policy breaks even, and in a manner that does not entail redistribution. The investor then makes an insurance decision followed by an investment decision, based on its (presumed accurate) information concerning the probability of reform, and taking into account the government's transition policy. Subsection 3 discusses the intuition behind the results and presents some of the most direct implications. The final subsection considers special cases -- risk neutrality and insurance where investment levels or states are observable -- in which a first best is achieved in the absence of government relief. These cases reinforce the intuition of the base case model, where moral hazard is the focus, and are of significance in their own right because they are directly relevant in a large variety of transition contexts.

2.1 Framework and Assumptions

2.1.1 Optimal Government Policy

This paper draws a distinction between new government policies and the policy concerning how transitions to such new policies are to be made, although some of the possible interrelationships will be explored in section 4. Until then, the argument will assume that government policies (other than transition policy) are optimal at every point, in terms of the information then available.[7] Reforms therefore arise in response to changing circumstances or new information. This may appear unrealistic in many contexts, depending upon one's faith in government and beliefs about biases in the political process. For the purpose of an initial normative analysis, however, it seems useful to analyze optimal transition policy in the context of a decisionmaker attempting to make optimal decisions generally. In some cases this might be a good approximation; when advising government officials, a contrary framework may not be feasible; and the ideal case is a useful point of departure for considering various second best contexts. If one makes the opposite assumption that the reforms themselves are undesirable, then it is rather obvious that transition policies which delay or undermine the impact of the policies to the maximum extent possible would be desirable. It should be noted that *some* particular assumption concerning the underlying substantive policy is necessary because a baseline of social benefits is necessary to evaluate efficiency properties of transition relief that take into account the effect of such relief on investment decisions.

Two related assumptions are also used. First, it is assumed that reforms are not motivated by distributional or revenue-raising goals; if they are, compensation would simply nullify the reform, and further analysis would be

----

[7] More precisely, in the notation of the model, the benefit functions, $B_i(K)$, measure both private and social benefits. The assumption in subsection 2.1.3 that there are not additional market imperfections is also necessary. An important effect of these combined assumptions is to rule out second best considerations.

unnecessary. Second, the effects of unmitigated government reforms do not reflect accounting or related changes. For example, if the government changed tax reporting periods (e.g., from calendar year to July-June fiscal year), it would be appropriate to provide a transition adjustment involving the one-time taxation of the half-year (January-June), rather than leaving those six months untaxed or double-taxed, depending upon the effective date of the reform. It will be argued in subsection 5.4.1 that many issues that have been discussed concerning the transition from an income to a consumption tax are of this character.

2.1.2 Consistent Transition Policy

There is a difference between whether compensation of those adversely affected by a government reform is optimal ex post and whether ex ante the government should have adopted a general policy of compensating various categories of losers from policy change. A government's policy toward transitions will be anticipated and thus will affect behavior, and individual transition decisions will presumably influence perceptions of the government's general policy. A static analysis that simply compares the states that would result from each version of the reform would ignore the effect of transition schemes upon behavior.

This study addresses what transition policy would be optimal assuming that it was to be followed consistently and its stability was credible to all actors. Uncertainty or mistaken perceptions concerning the transition policy itself would alter ex ante incentive effects and risk allocation. Subsection 3.3 will briefly consider this issue. Moreover, a different policy would often be optimal ex post so long as it was not expected ex ante and had no effect on expectations concerning future transition policies. Since the latter conditions are difficult to sustain in the long run, it seems more useful to determine the optimal consistent transition policy. Some remarks concerning the credibility of such a policy and its applicability in the short run before it is established will be offered in subsection 4.6.

2.1.3 Market Imperfections

For the most part, I will assume that financial markets have no imperfections except for moral hazard (including equivalent agency) problems. Many other imperfections would have no affect on the analysis so long as they implied no obvious remedy available to the government, and even those the government could remedy would often not alter the conclusions because direct remedies rather than modification of transition policy often would be the best solution. (The same assumption is made for other imperfections, except for those explicitly the target of the government policies being analyzed.) Of particular relevance in some contexts, most of the discussion here ignores transaction costs, misperception of risk, and adverse selection, all of which will be taken up in section 4.

2.1.4 Efficiency versus Other Considerations

Finally, some differences between this study and most past examinations of these problems should be noted. Gains and losses imposed by government reforms have generally been discussed under the rubric of horizontal equity -- the concern for equal treatment of equals, or, as it often is characterized, for preserving the pre-reform distribution of income or utility. Most investigations, such as Brennan (1971), King (1983), Plotnick (1981, 1982, 1984), and Rosen (1978), have focused on how the adverse impact of these distributional effects should be measured, which would presumably be useful in determining whether a proposed reform was desirable. This study, by contrast, focuses on mechanisms that might be employed to mitigate these effects. In addition, prior examinations of horizontal equity treated the concept as something to be valued for its own sake. Kaplow (1985a) criticized such formulations and argued that conventional risk analysis offered a more persuasive characterization of the concern that seems to motivate our instinctive revulsion to the arbitrary incidence of gains and losses.

Therefore, this study will proceed exclusively in those terms, which has the benefit that government-created risk and market risk can be compared directly and that risk and other efficiency considerations can be more readily combined and weighed. To the extent that the argument developed here concerning the close connection between government and market risk is convincing, there would be additional support for the view that horizontal equity concerns are best reinterpreted as reflecting a concern for risk.

This paper also will not address more conventional distributional concerns (vertical equity), except for a brief discussion in subsection 4.9. In general, such concerns are often better addressed directly, rather than through myriad modifications of other policies. Moreover, since it is highly implausible that most transition losses will be closely correlated with general inequality in the income distribution, transition relief typically will not be a fruitful channel to address this issue in any event.

2.2 Base Case Model

2.2.1 Structure of the Model

In the first period, individual investors make decisions, selecting a single level of investment. Returns are realized in period two. (The cost of investment and any returns in period one can simply be incorporated into the period two net returns, and discounting can be deemed implicit as well.) There is uncertainty as to the state of nature in period two. For simplicity, there are two states, one corresponding to that in which there is a government reform and the other being that in which there is not.[8] One can think of the

----

[8] The state of "no reform" is in many respects conceptually arbitrary. All that is necessary is for the two states to be fully specified, and, for the purposes of determining compensation (or windfall taxation), one be designated the reference state, or status quo. More on this in subsection 2.3.2.

In addition, the use of two states is artificial, in that it suggests that the insurance company would be able to make sufficient inferences from its ex net benefits from a given investment being affected solely by the government action, although the model is sufficiently general that changes in the underlying state of nature (independent of the government's reaction) could be a partial (or complete) source of the difference in benefits between the two states.

In period one, the government announces a compensation policy, which consists of a promise (that is believed and will be carried out) to pay a given percentage of the loss in period two in the event of "reform." Note that the loss can be negative (a gain) in which case the payment is negative (a "windfall tax"). Because of familiarity, the discussion will speak in terms of reforms generating losses, that are compensated, unless otherwise specified. The government also announces a lump sum tax, or "compensation premium" (in the event of transition gains, this would be a lump sum subsidy), which is equal to the expected value of compensation in period two. (For ease of notation, this tax is assessed in period two.) The motivation for including such a tax term -- which provides that, on an expected value basis, the government will break even -- is that the purpose of the model is to assess the welfare properties of compensation, not redistribution. Equivalently, one can compare a policy of compensation to a policy that transfers the same amount to the investor on an expected value basis, but as a certain, state-independent transfer. The analysis is clearer using the former approach, so that different compensation programs can be compared in a manner that involves the same aggregate transfer to the individual. (Without this addition, more compensation would necessarily increase expected utility, a rather uninteresting result that fails to get at the potential benefits or costs or transition relief.)

----

post observations to eliminate moral hazard. Subsection 2.4.2 discusses these possibilities, and suggests that moral hazard will typically be problematic for private insurance only when there are multiple states. The approach here is to use only two states for expositional simplicity, while simply assuming that the insurance company acts as if it does not have sufficient information to combat moral hazard. The case of multiple states is presented briefly in subsection 2.3.3, which indicates that there is little effect on the interpretation of the results. A more complicated modification appears in subsection 3.4

After the government announces its compensation policy (including the lump sum tax), the individual purchases insurance, which involves coverage for a given percentage of the loss (the individual may choose any amount of coverage) and an actuarially fair premium (collected in period two, for notational simplicity). Finally, the investor makes an investment decision. The insurance company sets the premium with the awareness that it is unable to observe either the state or the level of investment.[9] These assumptions are often artificial, and will be relaxed in subsection 4. Since they make moral hazard most intractable, insurance is most problematic, and thus compensation should be most likely to be helpful in spreading risk.

The notation is as follows:

U: investor's utility function (the model has only one investor); $U' > 0$, $U'' < 0$

K: amount of investment

$B_i(k)$: net benefits from investment in state i; the two states are denoted by "r" for reform and "o" for the "null" state, or status quo; $B_i'' < 0$, i=o,r; $\nexists$ K such that $B_i'(K)=0$, i=o,r;[10] $\Delta B \neq 0$,[11] where $\Delta B \equiv B_o - B_r$ -- "Δ" notation will have the analogous meaning for derivatives of *B*

Y_i: investors net income in state *i*

a: private insurance premium

----

[9] It does have knowledge of the government's compensation scheme, and premiums can be made a function of the total private coverage that is purchased. This two-stage model of moral hazard has some similarity to the model in Pauly (1974) involving multiple providers of insurance, each assumed to be unable to make its premium a function of the level of coverage purchased through other insurers. In the model here, the government is in some respects like the insurers in Pauly's model, although the insurance companies here are not.

[10] Given the stated conditions on the second derivatives, if such a K existed, it would be the optimum and there would be no moral hazard.

[11] If the optimal investment (K) given no insurance (b=0) is such that $\Delta B=0$, then the optimum entails no insurance. In this simple case, there is no risk to insure against in the first instance.

b: portion of loss $(B_o - B_i)$ covered by private insurance in state *i*

c: lump sum tax, or "compensation premium"

d: portion of loss $(B_o - B_i)$ covered by government compensation in state *i*

p: probability of state *r*

The government's problem is:

$\max_{d} U$, subject to:

(1) $c = pd\Delta B$

(2) private maximization (optimal insurance decision):

$\max_{b} U|_{d}$, subject to:

(i) $a = pb\Delta B$

(ii) investor maximization, which consists of:

$\max_{K} U|_{d,b}$

Note that the utility function is simply:

(3) $U = (1-p)U_o + pU_r$

$= (1-p)U[B_o - a - c] + pU[B_r - a - c + (b+d)\Delta B]$

2.2.2 Analysis of the Model

Starting with (ii), we can determine:

(4) $\frac{\partial U}{\partial K} = (1-p)U_o'B_o' + pU_r' \cdot [B_r' + (b+d)\Delta B'] = 0$

To determine $\partial U/\partial b$ from (2), one can substitute for *a*, using constraint (i), and keep in mind that *K* is a function of *b*, whereas (2) has *d* taken as given.

$$(5)\quad \frac{\partial U}{\partial b} = (1-p)U_o' \cdot [B_o'K_b - pb\Delta B'K_b - p\Delta B]$$

$$+ pU_r' \cdot [B_r'K_b - pb\Delta B'K_b - p\Delta B + (b+d)\Delta B'K_b + \Delta B] = 0$$

The sum of the first U_o' term and the first and fourth U_r' terms equals zero, based on $\partial U/\partial K$ (4).[12] Divide all remaining terms by p,[13] and group as follows:

$$(6)\quad -b\left[(1-p)U_o'\Delta B'K_b + pU_r'\Delta B'K_b\right] - \Delta B\left[(1-p)U_o' - (1-p)U_r'\right] = 0$$

Define

$$(7)\quad \bar{U}' \equiv (1-p)U_o' + pU_r',$$

so we now have:

$$(8)\quad (1-p)\Delta B(U_r' - U_o') - b\Delta B'K_b\bar{U}' = 0, \text{ or}$$

$$(9)\quad b = \frac{(1-p)\Delta B(U_r' - U_o')}{\bar{U}'\Delta B'K_b}$$

This result suggests properties of *b*, the percentage of insurance coverage, that are independent of *d*, the percentage of compensation, except to the extent that compensation affects the total amount of income in each state, as it surely will. The interpretation is as follows: ΔB is positive (in the hypothesized example, there is a loss due to the reform).[14] $U_r' - U_o'$ is

----

[12] As explored in Grossman and Hart(1983), for this substitution necessarily to be valid, the solution to (4) must be a unique maximum. The second order condition (16) (see page 22) is strictly negative for $b+d \in [0,1]$, which covers most cases of interest. Also, many of the proofs do not depend on the first order condition for *b*. For example, consider the four base case propositions. In Proposition I, it is only used to rule out b=0 as an optimum, which could only be problematic when $d \notin [0,1]$. Lemma II.1 also does not use the first order condition for *b*, and, given Proposition I(a), the sufficient condition that $b+d \in [0,1]$ holds for the remainder of the proof in Proposition II. Proposition III similarly does not use the first order condition for *b*, and Proposition IV is by assumption (plus Proposition I(a)) confined to a domain where the problem cannot arise.

[13] Later references in this subsection that examine $\partial U/\partial b$ when it is not set equal to zero will also have the "p" dropped from the equation, as it has no effect on any of the analysis or interpretations.

[14] If ΔB is negative -- i.e., if the reform results in a gain rather than a loss -- then $U_r' - U_o'$ is negative under the specified condition, so the numerator is still positive. The interpretation now is that the insurance company collects part of the uncertain gain, and the "premium" (non-state-contingent transfer) represents a payment to the insured.

positive, so long as $b+d < 1$. (If total coverage equals 1, income is equalized across states, and thus the marginal utilities are equal as well. If $b+d > 1$, there is "excess" coverage in that the resulting income is higher in the adverse state, making the term negative.) $\bar{U}'$ is necessarily greater than zero. $\Delta B' K_b$ being positive is thus consistent with $b > 0$. The interpretation of this final relationship is that $\Delta B' > 0$ implies greater marginal benefits in the state with greater total benefits (a relationship that will generally be helpful in interpreting the results). For the denominator to be positive in this instance, it must be that $K_b > 0$, which indicates that greater coverage increases investment. Likewise, if there were greater marginal returns in the adverse state, greater coverage, by placing more weight on the marginal returns in the favorable state, which are now assumed to be lower, suggests that investment would decline.

The first proposition characterizes how government compensation affects the private insurance decision. For compensation that is incomplete, one expects positive private insurance because, at the margin, when $b=0$, increasing *b* reduces risk, and the adverse incentive effect is borne by the government. Total coverage (private insurance plus compensation) will be less than 100%, because at that point there are no risk-spreading benefits at the margin but positive incentive costs. Similarly, if full compensation is offered, one expects $b=0$, because there is no risk remaining to insure and any incentive costs are totally borne by the government. The case of $d > 1$ is symmetric to that of $d < 1$, following the same reasoning.

Proposition I:

(a) $d < 1 \Rightarrow b > 0$ and $b+d < 1$.

(b) $d=1 \Rightarrow b=0$.

(c) $d > 1 \Rightarrow b < 0$ and $b+d > 1$.

Proof that $d < 1 \Rightarrow b > 0$: Reconsider the first order condition for maximization over *b* (8):

$$(8)\quad \frac{\partial U}{\partial b} = (1-p)\Delta B(U_r' - U_o') - b\Delta B'K_b\bar{U}' = 0$$

For b < (1-d), b+d < 1, which implies that the first term is positive: First, note that $Y_r - Y_o = (b+d-1)\Delta B$. Therefore, $\Delta B>0$ implies that the difference in marginal utilities is positive, and $\Delta B<0$ implies that it is negative. At b=0, (8) is positive.

For $d \le 1$, consider $b < 0$ as a possible optimum. Let *K* be the optimal investment given *b*. Compare the utility level thus resulting with the utility produced by the same *K* and b=0. The difference in expected incomes between the two cases is simply $-a + pb\Delta B$, which equals zero given (2: i). But for $d \le 1$ and $b < 0$, the spread in incomes is greater than with b=0, so expected utility must be less. Therefore, b < 0 also cannot be an optimum. *Q.E.D.*

<u>Proof that $d < 1 \Rightarrow b+d < 1$</u>: If b+d=1, $Y_o = Y_r$, so the first term in (8) equals zero. The sign of the second term can be determined from the following lemma:

<u>Lemma I.1</u>: *$b+d=1 \Rightarrow \Delta B'K_b > 0$.*

<u>Proof</u>: Begin with the first order condition for optimization over K:

$$(4)\quad \frac{\partial U}{\partial K} = (1-p)U_o'B_o' + pU_r'\cdot[B_r' + (b+d)\Delta B'] = 0$$

Take the derivative with respect to b:

$$(10)\quad \frac{\partial^2 U}{\partial b \partial K} = (1-p)U_o'B_o''K_b + (1-p)B_o'U_o''\cdot[B_o'K_b - pb\Delta B'K_b - p\Delta B] + pU_r'\cdot[B_r''K_b + (b+d)\Delta B''K_b + \Delta B']$$
$$+ p[B_r' + (b+d)\Delta B']U_r''\cdot[B_r'K_b + (b+d-pb)\Delta B'K_b + (1-p)\Delta B] = 0$$

To determine the sign of K_b, group the terms as follows:

$$(11)\ K_b\Big[(1-p)U_o'B_o'' + (1-p)B_o'U_o''\cdot[B_o'-pb\Delta B'] + pU_r'[B_r''+(b+d)\Delta B''] + $$
$$+ p[B_r'+(b+d)\Delta B']U_r''\cdot[B_r'+(b+d-pb)\Delta B']\Big]$$
$$= p\Delta B(1-p)B_o'U_o'' - pU_r'\Delta B' - p\Delta B(1-p)[B_r'+(b+d)\Delta B']U_r''$$

From (4), it can be seen that b+d=1 implies $B_o' = 0$: The bracketed portion of the second term in (4) is

$$(12)\ B_r' + (b+d)\Delta B' = B_r' + 1\cdot(B_o' - B_r') = B_o'$$

Therefore, (4) reduces to $\bar{U}'B_o'=0$. The right side terms in (11) are zero except for the second, which is strictly negative in the case where $\Delta B' > 0$, and the second and fourth terms on the left side equal zero, with the other terms strictly negative, so $K_b > 0$. And $\Delta B' < 0$ makes the right side positive without affecting the left, so $K_b < 0$.[15] *Q.E.D.*

If b≠0, b+d=1 is thus not an optimum, and a smaller *b* is locally preferred. And b=0, with b+d=1, contradicts that d < 1.

Consider b+d > 1; i.e., b > 1-d. Let $\hat{K}$ be the optimal investment under that scheme, and $\tilde{K}$ be the optimal investment when b+d=1. (Recall from (12) that (4) ⇒ $B_o'(\tilde{K})=0$.) In (4), the bracketed part of the second term equals $B_o' + (b+d-1)\Delta B'$. For the case where $\Delta B' > 0$,[16] since an optimum requires (4) to equal zero, it must be that $B_o' < 0$. By assumption, $B_o'' < 0$. Therefore, $\hat{K} > \tilde{K}$. Compare the utility level at *b* and $\hat{K}$ with that at b=1-d and $\hat{K}$. $\hat{K} > \tilde{K}$ implies that expected income is greater at b=1-d, because all B's are the same (since the same *K* is assumed) and the change in the insurance policy produces a gain in expected income because the cost per unit of remaining coverage is less (since the level of investment associated with b=1-d is less). Moreover,

----

[15] The possibility that $\Delta B'=0$ needs no special consideration, as it is clear from (4) that this would only occur if $B_o' = B_r' = 0$, which was ruled out by assumption.

[16] For the case where $\Delta B' < 0$, the relationship between the levels of investment discussed in the text to follow is reversed, but this also produces greater expected income in the case where b=1-d do to reduced premiums (the combination of the greater *K* and the reversed sign of $\Delta B'$ produces the same effect as the lower *K* and $\Delta B' > 0$). The remainder of the argument in text thus holds in this case as well.

the greater expected income is distributed perfectly equally when b=1-d. These two effects imply that the utility level is greater as well, which rules out the possibility that $b > 1-d$ can be an optimum. *Q.E.D.*

Proof that $d=1 \Rightarrow b=0$: At d=1, b=0 implies that (8) equals zero (as both terms equal zero). $b < 0$ was ruled out above. $b > 0$ is ruled out by the argument just presented that $b > 1-d$ cannot be an optimum, since that argument holds for d=1. Therefore, at d=1, b=0 is the optimum. *Q.E.D.*

Proof that $d > 1 \Rightarrow b < 0$: This proof parallels that for $d < 1 \Rightarrow b > 0$. Assume b=0. Since $d > 1$, $b+d > 1$, so the first term in (8) is negative. The second term is zero, so (8) is negative, and this cannot be an optimum. (Locally, a smaller, i.e., negative, *b* would be preferred.)

Now consider $b > 0$ as a possible optimum. Let *K* be the optimal investment given *b*. Compare the utility level thus resulting with the utility produced by the same *K* and b=0. The difference in expected incomes between the two cases is simply $-a + pb\Delta B$, which equals zero given (2: i). But for $d > 1$ and $b > 0$, the spread in incomes is greater than with b=0, so expected utility must be less. Therefore, $b > 0$ cannot be an optimum. *Q.E.D.*

Proof that $d > 1 \Rightarrow b+d > 1$: This proof parallels that for $d < 1 \Rightarrow b+d < 1$. b+d=1 implies, as discussed before, that the first term in (8) equals zero. Since $d > 1$, $b < 0$, so Lemma I.1 implies that the first order condition is positive, contradicting that it can be an optimum, and suggesting that a larger *b* is locally preferred.

Consider $b+d < 1$; i.e., $b < 1-d$. Let $\hat{K}$ be the optimal investment under that scheme, and $\tilde{K}$ be the optimal investment when b+d=1. (Recall from (12) that $(4) \Rightarrow B_o'(\tilde{K})=0$.) In (4), the bracketed part of the second term equals $B_o' + (b+d-1)\Delta B'$. For the case where $\Delta B' > 0$,[17] since an optimum requires (4) to equal zero, it must be that $B_o' > 0$. By assumption, $B_o'' < 0$. Therefore, $\hat{K} < \tilde{K}$. Compare the utility level at *b* and $\hat{K}$ with that at b=1-d and $\hat{K}$. $\hat{K} < \tilde{K}$

----

[17] The argument follows, mutatis mutandis, for the case $\Delta B' < 0$, for the same reasons indicated in note 16.

implies that expected income is greater at b=1-d, because all B's are the same (since the same *K* is assumed) and the change in the insurance policy produces a gain in expected income. This is because *b* is negative, so the subsidy per unit of remaining coverage is greater (since the level of investment associated with b=1-d is greater). Moreover, the greater expected income is distributed perfectly equally when b=1-d. These two effects imply that the utility level is greater as well, which rules out the possibility that $b < 1-d$ can be an optimum. *Q.E.D.*

The next proposition establishes the general undesirability of transition relief. That transition relief is at best unnecessary readily follows from the fact that private insurance can mimic government compensation: i.e., aggregate coverage (b+d) resulting from any *d* can be matched by private insurance alone. (See Lemma II.1.) At d=0, the private maximization problem described in (2) is essentially the same as the government's, but with one less constraint. The effect of the added stage in the moral hazard problem when $d \neq 0$ is that private insurance decisions are distorted, which also produces a distortion in the investment decision.

Proposition II: *d=0 is the* unique *global optimum.*

Proof: For many purposes, it is useful to rewrite the maximizing insurance decision in a manner suggesting that the individual chooses total coverage (including compensation), even though the premium will only cover the portion paid by the insurance company.

(13) $\beta \equiv b+d$, which implies $a = p(\beta-d)\Delta B$

The first order condition for $\partial U/\partial K$ is unchanged, except that b is replaced by $(\beta-d)$. (8) becomes:

(14) $\frac{\partial U}{\partial \beta} = (1-p)\Delta B(U_r' - U_o') - (\beta-d)\Delta B' K_\beta \bar{U}' = 0$

Lemma II.1: *For any d,* $U|_{d,b} = \hat{U}$, *where* $\hat{U} \equiv U|_{0,d+b}$ *(and* b *refers to the optimal private insurance, given* d*). Moreover,* $K=\hat{K}$.

Proof: Assume K maximizes $U|_{d,B}$. For d=0, the insurance company considers premium $\hat{a} = p(d+b)\Delta B(K)$ for coverage of d+b, which will be correct if K maximizes $\hat{U}$. Noting that $\hat{c}=0$ (since $\hat{d}=0$), one has:

(15) $\hat{a} + \hat{c} = pd\Delta B(K) + pb\Delta B(K) = c + a$

From (3) and (4), it is therefore clear that K also maximizes $\hat{U}$, since the only other possible difference in the maximization is in the b+d term, which in the maximization of U is b+d and in the maximization of $\hat{U}$ is (b+d)+0. From (3), $Y_i = \hat{Y}_i$, so $U = \hat{U}$. *Q.E.D.*

Lemma II.1 immediately implies that d=0 is a global optimum, since the utility achievable at any $d \neq 0$ can also be achieved at d=0. To prove it unique, consider an optimum where $d \neq 0$. At that optimum, the first order condition for β must hold (i.e., (14) must equal zero). By Lemma II.1, (14) also must equal zero for ($\tilde{d}=0$, $\tilde{b} \equiv d+b|_d$), since *d* is assumed to be a global optimum. Lemma II.1 implies $U=\tilde{U}$ and $K=\tilde{K}$, so (14) can equal zero in both cases only if $bK_b = \tilde{b}\tilde{K}_b$. (Note that $K_b = K_\beta$, when β=b+d and both are evaluated at the same *d*.)

The previous derivation concerning K_b will allow us to rule out this possibility. Begin by recalling:

$$K_b\Big[(1-p)U_o'B_o'' + (1-p)B_o'U_o''\cdot[B_o'-pb\Delta B'] + pU_r'[B_r''+(b+d)\Delta B''] + p[B_r'+(b+d)\Delta B']U_r''\cdot[B_r'+(b+d-pb)\Delta B']\Big] = p\Delta B(1-p)B_o'U_o'' - pU_r'\Delta B' - p\Delta B(1-p)[B_r'+(b+d)\Delta B']U_r'' \quad (11)$$

It simplifies matters and aids interpretation, however, to begin by performing a substitution on the left side, using the second order condition for maximization over *K*:

$$(16)\quad \frac{\partial^2 U}{\partial K^2} = (1-p)U_o'B_o'' + (1-p)B_o'^2U_o'' + pU_r'[B_r''+(b+d)\Delta B''] + p[B_r' + (b+d)\Delta B']^2U_r'' \leq 0$$

The bracketed term on the left side of (11) thus becomes:

$$(17)\quad \frac{\partial^2 U}{\partial K^2} - pb\Delta B'\left[(1-p)B_o'U_o'' + p[B_r' + (b+d)\Delta B']U_r''\right]$$

Using this result, (11) can be rewritten as follows:

$$(18)\quad K_b[U_{kk}-b\theta] = \Omega$$

Note that θ, Ω, and U_{kk} each depend only on β, U, and K. Lemma II.1 therefore guarantees that these three terms are identical in both cases. Therefore, we can equate

$$(19)\quad K_b[U_{kk}-b\theta] = \tilde{K}_b[U_{kk}-\tilde{b}\theta]$$

Factoring, we have:

$$(20)\quad bK_b[U_{kk}/b - \theta] = \tilde{b}\tilde{K}_b[U_{kk}/\tilde{b} - \theta]$$

From the requirement that $bK_b = \tilde{b}\tilde{K}_b$, we have the result that $b=\tilde{b}$. Recalling that $\tilde{b}\equiv d+b$, one gets $d=0$, contradicting that there can exist a globally optimal d not equal to zero.

This derivation implicitly assumed the following:

- $bK_b \neq 0$
- $U_{kk} \neq 0$
- bK_b and $\tilde{b}\tilde{K}_b$ are not both infinite

First, if $bK_b=0$, for (14) to equal zero, it must be that marginal, and thus total utility is equalized across states, which holds only if $\beta=1$. In that case, from Lemma II.1, the utility equals that at ($d=0$, $b=1$), which Proposition I demonstrates is not an optimum since $d=0 \Rightarrow b < 1$.

$U_{kk} = 0$ can be ruled out by inspection of (16) so long as $\beta \in [0, 1]$. Lemma II.1 and Proposition I guarantee that $\tilde{b} \in (0, 1)$ and $\tilde{b} \equiv \beta$, so at an optimal d, this holds.

Finally, given *d*, for bK_b and $\tilde{b}\tilde{K}_b$ both to be infinite, both K_b terms must be infinite. From (11), this requires $U_{kk} - b\theta = 0$ and $U_{kk} - \tilde{b}\theta = 0$. This is only possible if $b = \tilde{b}$, the desired conclusion, or $\theta = 0$ *and* $U_{kk} = 0$ simultaneously, the latter just having been ruled out. *Q.E.D.*

The remaining two propositions characterize the effect of $d \neq 0$ on the level of aggregate coverage, β. Following the previous intuition, one expects excessive aggregate coverage when $d > 0$, although private coverage considered alone would probably tend to be less. (And conversely for $d < 0$.) Because of a number of income effects, some additional restrictions are necessary to guarantee this precise characterization.

Proposition III: *If $\Delta B > 0$ and U is a nonincreasing absolute risk aversion utility function, then $d > 0 \Rightarrow \beta^*|_d > b^*$ and $d < 0 \Rightarrow \beta^*|_d < b^*$, where b^* is the optimum at $d=0$.*

Proof for $d \geq 1$: Proposition I demonstrates that $d \geq 1 \Rightarrow \beta \geq 1$ and $b^* < 1$. *Q.E.D.*

Proof that $\beta^*|_d \neq b^*$: Lemma II.1 implies that both would yield the same utility, which contradicts Proposition II's demonstration that $d=0$ is a *unique* global optimum. *Q.E.D.*

Proof that $d \in (0,1] \Rightarrow \beta^*|_d \not< b^*$: Assume the contrary. Consider the following four combinations:

- $U \equiv U|_{d,b}$, where $b \equiv b|_d$

- $U^* \equiv U|_{0,b^*}$

- $\hat{U} \equiv U|_{d,\hat{b}}$, where $\hat{b} \equiv b^* - d$

- $\tilde{U} \equiv U|_{0,\tilde{b}}$, where $\tilde{b} \equiv b$

Notation here will refer to the utility achieved when *K* is optimized, given the values of *c* and *a* that would correspond to each situation, except to the extent otherwise indicated. (E.g., $\hat{U}|_{K^*}$ refers to the utility achieved at (d,b*-d) and the values of *c* and *a* this scheme would entail, evaluated at K=K* instead of $K=\hat{K}$.[18])

The following Lemma (III.1) is necessary to prove the main Lemma (III.2).

<u>Lemma III.1</u>: *If $\Delta B > 0$, $\beta \in [0,1]$, $b \geq 0$, and* U *is a nonincreasing absolute risk aversion utility function, then $\Delta B' K_b > 0$.*[19]

<u>Proof</u>: Begin by recalling the expression for K_b:

$$(11)\quad K_b\Big[(1-p)U_o'B_o'' + (1-p)B_o'U_o''\cdot[B_o'-pb\Delta B'] + pU_r'[B_r''+(b+d)\Delta B''] + + p[B_r'+(b+d)\Delta B']U_r''\cdot[B_r'+(b+d-pb)\Delta B']\Big]$$

$$= p\Delta B(1-p)B_o'U_o'' - pU_r'\Delta B' - p\Delta B(1-p)[B_r'+(b+d)\Delta B']U_r''$$

and the simplification for the bracketed term on the left side of (11):

$$(17)\quad \frac{\partial^2 U}{\partial K^2} - pb\Delta B'\Big[(1-p)B_o'U_o'' + p[B_r' + (b+d)\Delta B']U_r''\Big]$$

To further illuminate the remaining bracketed term, rearrange the first order condition for *K* (4) as follows:

$$(21)\quad p[B_r' + (b+d)\Delta B'] = \frac{-U_o'(1-p)B_o'}{U_r'}$$

Use (21) to substitute in (17):

$$(22)\quad \frac{\partial^2 U}{\partial K^2} - pb\Delta B'(1-p)B_o'\Big[U_o'' - \frac{U_o'}{U_r'}U_r''\Big]$$

Factor out U_o' from the rightmost term:

----

[18] Note that the value of *c* would still be based on *b* rather than $\hat{b}$ (i.e., $\hat{c} = c$), since *c* is set before *b* is chosen.

[19] The possibility that $\Delta B'=0$ can be ignored because it implies that $B_o' = B_r'$. Optimization over *K* therefore implies, from (4), that both marginal benefits equal zero, which was ruled out by assumption.

$$(23) \quad \frac{\partial^2 U}{\partial K^2} - p(1-p)bB_o' \Delta B' U_o' \left[\frac{U_o''}{U_o'} - \frac{U_r''}{U_r'} \right]$$

The factors preceding the bracketed term are positive as a group. (Assume $b > 0$.[20] From (4) it is clear that B_o' and $\Delta B'$ have the same sign.[21]) The bracketed term is simply the negative of the difference between the coefficients of absolute risk aversion. Since Y_o is greater than Y_r,[22] from the condition that $\Delta B > 0$, it follows from the assumption of nonincreasing absolute risk aversion that the term is either zero (in which case the proof will still follow), or positive: since the risk aversion coefficient is greater at U_r, the difference in coefficients is negative, and the bracketed term is the negative of that difference. Since $\beta \in [0,1] \Rightarrow U_{kk} < 0$, (23) -- which equals the left side bracketed term in (11) -- is negative.

Now examine the right side of (11), first for the case $\Delta B' > 0$. All three terms on the right side are negative: In the first term, ΔB is positive, $\Delta B' > 0$ implies that B_o' is positive, and U_o'' is negative. In the second term, $\Delta B'$ and U_r' are positive, so the term is negative. In the third term, the component in brackets is negative (from (4), since (4) must equal zero) and the first term and U_r' are positive. U_r'' is also negative, so the term as a whole is negative. Therefore $K_b > 0$.

To complete the proof, verify that $\Delta B' < 0$ implies $K_b < 0$. Back in (11), all three terms on the right side are now positive, following the same logic as before, and the bracketed term on the left side (23) is still negative, as there are counteracting sign reversals in B_o' and $\Delta B'$. *Q.E.D.*

Before proceeding to Lemma III.2, note that the two additional conditions of Lemma III.1 ($\beta \in [0,1]$ and $b \geq 0$) are satisfied for all cases now under

----

20 b=0 makes the whole term zero, which is sufficient for (23) to be negative.

21 At β=1, $B_o' = 0$ (12), which is sufficient for (23) to be negative. See also Lemma I.1.

22 If β=1, income is equal, and the comments in the preceding note apply.

consideration, as demonstrated by the application of Proposition I to $d \in (0,1]$.

Lemma III.2: *Given the conditions of Proposition IV and $d \in (0,1]$, $\beta^*|_d \Rightarrow U \leq \hat{U}$.*

Proof: Compare U^* and $\hat{U}|_{K^*,a=\alpha}$ -- i.e., utility given d, c, $b \equiv b^*-d$, $K=K^*$, setting $a=\alpha$ -- where $\alpha \equiv a^*-\hat{c}$. Since β and K are the same in both cases, and a+c is also equal in both cases ($a^* + c^* = a^* + 0 = a^*$; $\alpha + \hat{c} = a^* - \hat{c} + \hat{c} = a^*$), utility will be equal.

For the case of $\Delta B' > 0$, Lemma III.1 implies $K^* > \tilde{K}$, and Lemma II.1 implies $\tilde{K}=K$. Therefore, $K^* > K$, which implies:

$$(24)\ \hat{c} = c = pd\Delta B(K) < pd\Delta B(K^*)$$

Also, $a^* \equiv pb^*\Delta B(K^*)$, so we have

$$(25)\ \alpha = a^* - \hat{c} > p(b^*-d)\Delta B(K^*)$$

If $K^* \geq \hat{K}$, then $\alpha > \hat{a}$, which implies $U^* < \hat{U}|_{K^*,a=\hat{a}} \leq \hat{U}$.

Consider the possibility that $\hat{K} > K^*$. If it is still the case that $\alpha \geq \hat{a}$ (which is possible), the result still follows (although not necessarily with strict inequality). Assume $\hat{a} > \alpha$. By construction, at $(d,\hat{b})$, α induces K^* (same a+c and β as at U^*), and by definition $\hat{a}$ induces $\hat{K}$. This configuration implies that in part of the interval $[\alpha,\hat{a}]$, $K_a|_{\hat{b}} \equiv \partial K/\partial a|_{\hat{b}} > 0$. This possibility can be ruled out. Begin by taking the derivative of the first order condition for *K* (4) with respect to *a*:

$$(26)\ \frac{\partial^2 U}{\partial a \partial K} = (1-p)U_o'B_o''K_a + (1-p)B_o'U_o''\cdot[B_o'K_a - 1] + pU_r'\cdot[B_r''K_a + (b+d)\Delta B''K_a]$$
$$+ p[B_r' + (b+d)\Delta B']U_r''\cdot[B_r'K_a - 1 + (b+d)\Delta B'K_a] = 0$$

Recalling the expression for U_{kk} from (16), we have:

$$(27)\ K_a U_{kk} = (1-p)B_o'U_o'' + p[B_r' + (b+d)\Delta B']U_r''$$

The right side is the bracketed term in (17) used in Lemma III.1, and, as the analysis there demonstrates, has the same sign as $\Delta B'$ (or equals zero, in the constant absolute risk aversion case), and U_{kk} is negative. Therefore, $K_a \le 0$ throughout, a contradiction.

To complete the Lemma, consider the case of $\Delta B' < 0$. Following the above analysis, it is clear that $K^* < K$, and K_a reverses sign. The argument thus follows, mutatis mutandis. (This time, $K^* \le \hat{K}$ uses the first argument concerning the level of α, and $K^* > \hat{K}$ used the second argument concerning the sign of K_a.) *Q.E.D.*

Lemma III.2 establishes $U^* \le \hat{U}$. Since *U* is the optimum, given *d*, it follows that $\hat{U} \le U$. Proposition II establishes that d=0 is the *unique* global optimum; thus, $U < U^*$, establishing a contradiction. *Q.E.D.*

Proof that $d < 0 \Rightarrow \beta^*|_d \not> b^*$: This result follows from the previous one, mutatis mutandis. One begins with the parallel construction. The inequality in (24) still holds: this time, $K^* < K$, which causes one reversal in the direction of the inequality, but $d < 0$ causes another. For the analysis concerning the sign of K_a, the only differences arises because it is possible that $U_{kk} = 0$ when $\beta < 0$, but in that event the proof would be complete since $b^* > 0$. ($\beta < 0$ also creates problems for the application of Lemma III.1, as discussed in the Appendix, but no problem for the claim of Proposition III.) *Q.E.D.*

Note on Risk Aversion Assumption in Lemma III.1: A brief word is in order concerning the intuition behind the assumption of nonincreasing absolute risk aversion and the sign of the second term on the left side of (11). The direct effect of an increase in *b* is to lower Y_o. This is reflected in the first term on the right side of (11). In the second term on the left side, the bracketed component -- $[B_o' - pb\Delta B']$ -- reflects the indirect effect on Y_o resulting from the change in *K* induced by the change in *b*. If that term is positive, then the second term overall is unambiguously negative (so long as $U'' < 0$). But if that term is negative, consider the implication of assuming that $K_b > 0$. The effect of an increase in *K* induced by the increase in *b* is to decrease Y_o beyond the direct decrease caused by greater coverage. If U_o'' is relatively "large" -- increasing absolute risk aversion, of course, is a relative statement concerning the magnitude of U_0'', captured in the combination of portions of the second and fourth terms (23) -- the direct decrease in Y_0 will already have produced a relatively large increase in U_0'. A sufficiently large increase would be inconsistent with the hypothesized desirability of further reducing Y_0 by increasing K in response to the increase in b, rather than increasing Y_0 (partially offsetting the direct reduction caused by the increase in b) by decreasing K. This suggests the possibility that $K_b < 0$ if absolute risk aversion is sufficiently greater at the higher income level, Y_0.

It is also interesting to note that $\partial K/\partial a$ (holding b fixed) equals the bracketed term in (17) divided by U_{kk}. (Recall (27) in Lemma III.2.) Therefore, when $\Delta B' > 0$, the nonincreasing absolute risk aversion assumption implies that $\partial K/\partial a \leq 0$. Since $\partial K/\partial a = -\partial K/\partial Y$, this implies $\partial K/\partial Y \geq 0$; i.e., that increasing income encourages investment when absolute risk aversion is decreasing with income. (And, when $\Delta B' < 0$, i.e., increased investment decreases the difference in benefits between the two states, the opposite result holds.)

<u>Proposition IV</u>: *If $\Delta B > 0$, $d \in (0,1]$, and* U *is a nonincreasing absolute risk aversion utility function, then:* $\partial U/\partial \beta|_{0,\beta} < 0$, *where* $\beta \equiv d + b|_d$.

<u>Proof:</u> In Appendix.

Remarks on the Case $\Delta B < 0$: Most directly, this corresponds to the case of transition ("windfall") gains, to be discussed further in subsection 2.3.2. As a formal matter, it is interesting to note where the previous arguments go awry when $\Delta B < 0$. Only Propositions III & IV depended on the contrary assumption, and only one aspect of either proof depended upon it: the argument of Lemma III.1 demonstrating that, under specified conditions, $\Delta B' K_b > 0$. As described earlier, this condition corresponds to an intuitive understanding of the implications of moral hazard -- e.g., when $\Delta B' > 0$, one would expect $K_b > 0$ because increasing K in response to an increase in b increases insurance payments in the adverse state more so than when b was lower. Moreover, the more equal spreading of income across states makes the investor less risk averse, also inducing an increase in K. (For this latter effect, note that in (4), an actuarially fair policy with a larger *b* results in an increase in U_o' and a decrease in U_r', the former state having positive marginal benefits and the latter having negative marginal benefits.)

$\Delta B < 0$ reverses this latter effect, and thus can disturb the results. This can be seen in two places in the derivation of Lemma III.1. First, in analyzing the right side of (11), the first and third terms (the U'' terms) are multiplied by ΔB. Thus, if ΔB has the opposite sign, it is possible for the sign of the right side to reverse. These terms correspond to this marginal utility of income effect.[23] The second term, which remains unchanged, is the direct investment incentive effect due to the change in *b*. Second, in determining the sign of (23), it was assumed that $Y_o > Y_r$. $\Delta B < 0$ reverses this result as well, making it possible that the sign of the bracketed term on the left side also changes.[24]

It is interesting to note that these two problems with Lemma III.1 were precisely the reason the assumption $\beta \leq 1$ was necessary. $\beta > 1$ changes the first and third right side terms because B_o' becomes negative (in the $\Delta B' > 0$ case) and the bracketed term, which corresponds to the U_r' coefficient in (4), also changes sign. Similarly, on the interpretation of (23) there are three sign reversals, changing the sign of the second term: $b < 0$ (see Proposition

----

[23] This effect would have some relevance even in determining the effect of a change in *b* holding *a* fixed, because, for example, an increase in *b* in the $\Delta B > 0$ case raises income in the reform state, thus lowering marginal utility in that state, which, when $\Delta B' > 0$, induces a further increase in *K* since marginal benefits are negative in that state. Conversely, for $\Delta B < 0$, this effect would cause *K* to fall, counteracting the direct incentive effect. This can be seen from the following expression:

$$\text{(N1)} \quad \frac{\partial K}{\partial b}\Big|_a = \frac{-pU_r'\Delta B' - p\Delta B[B_r' + (b+d)\Delta B']U_r''}{U_{kk}}$$

The denominator is generally negative, and, for the $\Delta B' > 0$ case, the first term is negative, and the second (including the preceding minus sign) is negative if and only if $\Delta B > 0$. The first can be seen as a substitution effect and the second as an income effect, the latter depending on the sign of ΔB.

[24] This discussion relates to the need for the nonincreasing absolute risk aversion assumption in Lemma III.1. That assumption was required to assure that risk aversion was less at U_o than at U_r. Either increasing absolute risk aversion or $\Delta B < 0$ causes a problem. (If both held, the effects counteract, restoring the determinate negative sign to (23).)

I); $B_o' < 0$, as just noted; and the bracketed term changes sign because $\beta > 1$ makes income higher in the reform state. It is no accident that $\beta > 1$ and $\Delta B < 0$ have such similar effects: both reverse the states in terms of which has greater income, and it was the income effects that were seen to be the source of the problem when $\Delta B < 0$.

$\Delta B < 0$ can also be interpreted as a redefinition of the status quo. If, in the $\Delta B > 0$ case, the states are re-named, one can perform a simple transformation of all the variables to make it the same problem, except that $\Delta B < 0$. Alternatively, one could redefine which state involves the payment of insurance and compensation (by redefining, for these purposes, $\Delta\tilde{B} \equiv B_r - B_o$). Once again, a $\Delta B > 0$ problem becomes a $\Delta B < 0$ problem. This discussion highlights the arbitrariness of the original status quo definition, reinforcing the suggestion to follow that transition gains and losses largely should be viewed symmetrically.

2.3 Discussion

One of the most significant limitations on the ability of private markets to spread all risks is that it is often necessary to have substantial risks borne by a limited group for incentive reasons (moral hazard). [See Arrow (1963), Holmstrom (1979), Jensen & Meckling (1976), Shavell (1979).] Thus, absent compensation, $b^* < 1$. Additional government insurance -- e.g., $d > b^*$, while forcing $b = 0$ -- would not offer a net improvement (and in fact would create a net detriment) for precisely the reasons that further private risk-spreading arrangements were found undesirable by the market. Private arrangements are optimal given the information available to insurers. [See Pauly (1968), Shavell (1982)]. So long as the government's information concerning optimal private behavior is not superior (see subsection 4.3), it can do no better.

The intuition behind the four propositions is that compensation, in a sense, creates an externality that did not otherwise exist. The reform affects the real (private and social) return on investment (B). Taking the reform as given, it is desirable for private actors to take its effects into account. This will happen appropriately when all the costs and benefits are directly borne, as is the result when d=0. That the private party might insure does not disturb this conclusion. Even though a first best is not achieved, there is no externality in the insurance decision, and therefore the optimal trade-off is achieved. When $d > 0$, the government bears part of the private costs, making them external to the private party. (When $d < 0$, excessive costs are imposed.) As a result, the private insurance and investment decisions are distorted. When $d > 0$, there is too much aggregate coverage (i.e., $\beta > b^*$). Note that $b > 0$ even when $d \geq b^*$ (so long as $d < 1$), again due to the "externality." Similarly, $\beta > b^*$ when $d \in (0,b^*)$ -- i.e., there is overshooting due to the "externality." Absent compensation, the investor bears the moral hazard cost because increases in *b* will be reflected in the premium charged. If $d > 0$, part of the moral hazard cost is borne by the government, so the increase in *a* is less than in the no compensation case. The only way to prevent this effect would be to prohibit insurance and set $d=b^*$ (or, equivalently, order that $\beta=b^*$, regardless of the *d* that is announced). Such a policy in essence amounts to compulsory government insurance of precisely the amount (and form) that the investor would have chosen in the absence of government relief (d=0).

As a simple example, assume that there is a 50% chance that the government will soon terminate a project that constitutes half the demand for widgets. A firm must decide how large a new factory should be build -- widget factories have large capital requirements that cannot be transferred to other uses. With full compensation, the adverse contingency is effectively ignored, resulting in inefficiently large capacity. Absent any relief, there still may be some excess investment, to the extent significant insurance is arranged, but any such contract will take this effect into account in setting the premium, producing a more efficient result, even taking risk into account.

2.3.1 Government and Market Risk Compared

The analysis presented thus far is equally applicable to "market" risk (i.e., any risk not caused by government policy change), where it has been more traditionally accepted by economists that government relief would generally be inefficient, for the sorts of reasons presented here. For concreteness, it is useful to consider some examples of changes in government policy and to compare them with parallel examples involving market risk.

1. *Pigouvian Tax vs. Changes in Factor Prices*: There is a probability (p_t) that in year *t* a Pigouvian tax (of known magnitude) will be imposed on each unit of a firm's output to correct an externality if and only if it also arises at time *t*.[25] Alternatively, assume there is the same probability that the costs of delivering the firm's finished product will increase by the same amount at time *t* -- e.g., due to increased fuel prices.

2. *Taking for Highway vs. Flood or Competition*: There is a probability (p_t) that in year *t* it will be desirable to build a highway, which will be routed through the property where a firm is currently contemplating how large a factory (if any) it should build.[26] Now assume, instead of probability (p_t) that a taking will occur at time *t*, that there is the same probability that a flood will destroy the investment at time *t*. Or that a competitor's new product will render a firm's product, and factory, worthless.

----

[25] The more familiar alternative where an external effect that has existed all along is discovered by a public authority at time *t* will be explored in subsection 3.2.

[26] See Blume, Rubinfeld, and Shapiro (1984), Blume and Rubinfeld (1984), Kaplow (1986). To simplify the picture, assume that any investment on the route of the highway must be destroyed, that the highway project is desirable regardless of the private investment decisions made along the route, and that no alternative routes are feasible. The first assumption is necessary to create a conflict in the first place, the second is often realistic, and the third is largely an extension of the second in that it only matters that an individual investment not affect which route is optimal.

3. *Government Demand vs. Private Demand*: There is a probability (p_t) that in year *t* the government's demand for widgets (or a particular type of engineers) will fall dramatically due to changes that eliminate the government's need for the product. Government demand is a substantial portion of total demand, and firms must make irreversible long-run investment decisions currently. Instead, consider the same prospect of a decrease in private demand (due to changing tastes, imports, etc.).

In each of these examples, the risks and incentive effects, as well as the welfare properties of requiring the firms to bear these risks, would be the same. In the above model, to make the formal application to market risk, one need only reinterpret the meaning of B_r.

2.3.2 Compensation for Losses Compared to Windfall Taxation of Gains

It is worth emphasizing that the same analysis used to evaluate a scheme providing for government compensation of losses provides parallel lessons concerning government taxation of windfall gains. Just as compensation for the imposition of a tax induces overinvestment ex ante, a one-time levy on the gain in value due to repealing a tax would lead ex ante to overdeterrence of the taxed activity.[27] For example, a new subsidy might be seen as providing a windfall gain to old investment, thus making a windfall tax or other transition provision appropriate. Yet, at the time the old investment was made, the prospect of the new subsidy (if unaccompanied by transitional mitigation) can be a significant part of the inducement for the investment. It is efficient for ex ante investment to respond to positive government prospects, so windfall taxation results in underinvestment. Similarly, since in both instances risk affects private values, risk-spreading considerations will be parallel.

----

[27] Each of the three previous examples has analogues involving gain. For Pigouvian taxes there are similar subsidies. Least straightforward is the taking since the government rarely gives away property or buildings, although provision of a public good has similar properties. And for decreases in government demand, there are increases as well.

One can simply reinterpret the model for the case where $\Delta B < 0$. (See note 14.) As discussed at pages 28-30, the formal results are largely applicable to this case, which can also be interpreted as a redefinition of the status quo in terms of the reference state from which compensation and insurance payments are measured. (See note 8 and page 30.) In areas such as tax policy, where significant reform is frequent, and often substantially anticipated, this latter point is all the more telling.

The connection between the analysis of compensation for losses and taxation of windfall gains is frequently overlooked in analysis of the former issue, as the vast majority of discussions argue about the extent of relief appropriate in the case of losses and never mention the parallel arguments for windfall taxation of gains.[28] Some windfall tax schemes have been considered or implemented. More prominent examples include proposals to tax increases in the land value of properties due to their proximity to new government projects and the taxation of windfall profits due to the lifting of energy price controls.[29] The previous arguments against compensation schemes are also arguments against such taxation schemes, and many of the arguments commonly offered in debates over such taxation schemes -- particularly adverse effects on incentives -- in fact parallel those developed here in evaluating compensation schemes.

It is not the case, however, that compensation and taxation schemes are the same in all relevant respects. The former require the government to pay out money whereas the latter collect revenues for the government. To the extent that government revenue-raising is costly, it would appear that compensation should be favored even less than the previous analysis dictates, whereas taxation schemes might be desirable despite their costs. In fact,

----

[28] For discussion of representative examples, see Kaplow (1986).

[29] Examples in the grandfathering context include determining whether liberalized depreciation policy or the investment tax credit will be available to old investment (with or without requiring a post-implementation sale), or the recent proposal concerning ACRS windfall recapture (motivated by a proposed reduction in tax rates). Some proposals to integrate the corporate income tax into the individual income tax would differentiate between new and old equity, to avoid the windfall gains as to the latter.

revenue-raising no doubt motivates many of the more prominent instances in which windfall taxation, or grandfathering in the case of gains, has been proposed. Outside the context of tax reform (where raising money is a central focus of attention to begin with) the costs associated with relief are less frequently considered, perhaps because in these other contexts the effects are more hidden, since monetary compensation or windfall taxation is not the typical form of relief used.

Of course, if revenue-raising were the only consideration, virtually anything could be taxed. The apparent attractiveness of transition relief is that "windfall" taxes are preferable because they are nondistortionary. The problem, however, is that to the extent gains are not true windfalls, there would be costs because of the incentive effects described here. It may be that such costs are less per unit of revenue raised than those associated with, for example, the income tax. The analogy is the familiar result that it is desirable to raise revenue (for a taxing authority, or a regulated monopoly that has a break-even constraint) by imposing greater taxes on activities subject to less distortion -- with commodity taxes, those with a lower elasticity of demand.

A complete evaluation of the revenue costs and benefits of transition relief would have to take into account distributional consequences, considered further in subsection 4.9. Consider two extreme possibilities. To the extent the incidence were largely random, one could compare a random head tax for raising revenue, which is generally opposed -- despite the fact it does not distort behavior -- because of its incidence. It might be thought that the windfall tax is a superior "random" device, because it dampens the amplitude of variations that otherwise occur, whereas the pure random head tax adds another source of variation.[30] The other extreme would be where the windfall

----

[30] This point relates directly to the connection between horizontal and vertical equity advanced in Kaplow (1985a). Randomness will generally lead to horizontal inequity. But viewing the ex post distribution, or ex ante *expected* utility, a random scheme that systematically dampened fluctuations would generally produce more vertical equity than one that added a further random component to an overall uncertain situation. Moreover, the problem discussed there concerning the arbitrariness of the definition of the status quo in motivations for horizontal equity has a direct analogue to that

tax scheme has a particularly favorable incidence. Suppose, for example, that the burden was closely correlated with income. This might be thought a stronger case for windfall taxation, but it must be taken into account that such correlations -- to the extent they can be counted on in the future -- are not happenstance, but rather must be a function of the same sorts of decisions that determine income levels. As a result, the greater is the correlation, the more one would expect the imposition of such a windfall tax scheme to have the incentive properties of the income tax that is being replaced. Any remaining random component could be analyzed along the lines of the first case. In general, one can compare a windfall tax policy to a tax scheme that raises similar revenue and has similar incidence. The difference will relate to the uncertain elements, which would appear more desirable for a windfall tax than for purely random components added to the latter,[31] were it not for the ex ante incentive effects of the windfall tax, described previously.

2.3.3 Multiple States

There typically are many possible future states. Moreover, the large number of possibilities is, as a practical matter, significantly responsible for the inability of private insurance to achieve a first best, as discussed in subsection 2.4.2. This subsection briefly restates the model for the case of more than two states. As one would expect, this does not fundamentally alter the results.

The following notation will be employed:

p_i: probability of state *i*

$\Delta B_i \equiv B_o - B_i$

----

discussed previously concerning interpretation of the sign of ΔB. I.e., once there exists uncertainty concerning future government policy, there is no unambiguous way to define a change in the "expected" policy, since that expectation directly incorporated a number of uncertain prospects.

[31] The caveat discussed in subsection 3.3, see page 88, concerning incentive effects of randomness in taxation, is relevant here.

$$Y_i \equiv B_i - a - c + \beta \Delta B_i$$

$$a = b\sum p_i \Delta B_i = b\Delta \overline{B}$$

$$c = d\sum p_i \Delta B_i = d\Delta \overline{B}$$

$$U \equiv \sum p_i U_i, \quad U_i \equiv U(Y_i)$$

The derivations directly parallel those in the simpler case involving two states.

$$(*1) \quad \frac{\partial U}{\partial K} = \sum p_i U_i'[B_i' + \beta \Delta B_i'] = 0$$

$$(*2) \quad \frac{\partial U}{\partial b} = \sum p_i U_i'[B_i' K_b + \beta \Delta B_i' K_b + \Delta B_i - b\sum p_i \Delta B_i' K_b - \sum p_i \Delta B_i] = 0$$

Note that $\Delta \overline{B}' \equiv \sum p_i \Delta B_i'$, $\Delta \overline{B} \equiv \sum p_i \Delta B_i$, and $\overline{U}' \equiv \sum p_i U_i'$. Making appropriate substitutions and canceling terms that (*1) indicates sum to zero yields:

$$(*3) \quad \frac{\partial U}{\partial b} = -b\Delta \overline{B}' K_b \overline{U}' + \sum p_i U_i'(\Delta B_i - \Delta \overline{B}) = 0, \text{ or}$$

$$(*4) \quad b = \frac{\Delta \overline{B} \sum p_i U_i'(\Delta B_i / \Delta \overline{B} - 1)}{\Delta \overline{B}' K_b \overline{U}'}$$

Each of these expressions has direct analogues in (4), (5), (8), and (9) respectively.

2.3.3.1 Applicability of Proposition I

Ruling out b=0: (*3) corresponds to (8), and to rule out b=0, it must be established that

$$(*5) \quad \sum p_i U_i'(\Delta B_i - \Delta \overline{B}) \neq 0.$$

In fact, for $d < 1$, it can readily be verified that this expression is positive, following the argument in Proposition I. Here, $\Delta \overline{B} \equiv \sum p_i \Delta B_i$. The greater the difference in parentheses, the lower is Y_i, and thus the greater is U_i'. Thus, the sum is positive.

Demonstrating $d \leq 1 \Rightarrow b \geq 0$: The argument of Proposition I holds directly.

Ruling out β=1 when $d \neq 0$: The analogue to Lemma I.1 -- demonstrating that $\beta=1 \Rightarrow \Delta B'K_b > 0$ -- can be demonstrated for this case by taking the derivative of (*1) with respect to *b*:

$$(*6)\ \frac{\partial^2 U}{\partial b \partial K} = \sum p_i U_i'[B_i''K_b + \beta \Delta B_i''K_b + \Delta B_i']$$
$$+ \sum p_i U_i''[B_i' + \beta \Delta B_i'][B_i'K_b + \beta \Delta B_i'K_b + (\Delta B_i - \Delta \bar{B}) - b\Delta \bar{B}'K_b] = 0$$

One can group terms and substitute for U_{kk} as follows:

$$(*7)\ K_b[U_{kk} - b\Delta \bar{B}'\sum p_i U_i''(B_i' + \beta \Delta B_i')] = -\sum p_i U_i' \Delta B_i' - \Delta \bar{B} \sum p_i U_i''(\frac{\Delta B_i}{\Delta \bar{B}} - 1)(B_i' - \beta \Delta B_i')$$

From (*1), it is clear as was the case before (12) that β=1 implies that $B_o'=0$. The left side of (*7) therefore reduces to K_bU_{kk}. Similarly, all the terms in the summation in the second right side term are multiplied by zero, leaving only the first term, which at β=1 is simply $-U'\Delta\bar{B}$. At $\beta=1$, $U_{kk}=\bar{U}'B_o''$, which is strictly negative, establishing that, in (*3), the first term is negative and the second is zero. (And, as before, (*3) holds if *b* also equals zero -- i.e., at d=1.) The only problem arises if $\Delta\bar{B}'=0$. As in the two-state case, if this holds at precisely the *K* that makes $B_o'=0$, there is an optimum at b=1 involving no moral hazard.

Demonstration that $d \leq 1 \Rightarrow \beta \leq 1$: Unlike the other arguments, this part of the proof runs into some difficulty. Without loss of generality, consider the case where $\Delta\bar{B}' > 0$. In the two-state case, it was demonstrated from the first order condition for *K* (4) that this implied, for $\beta > 1$, that $B_o' < 0$, and thus $K|_{\beta>1} > K|_{\beta=1}$. In this case, at $\beta > 1$, (*1) reduces to:

$$(*8)\quad \bar{U}'B_o' + (\beta-1)\sum p_i U_i' \Delta B_i' = 0$$

It can be demonstrated that B_o' is negative if it can be shown that the latter term is positive. $\Delta\bar{B} > 0$ is insufficient, because it is still possible that some of the $\Delta B_i'$ terms are negative, and, if B_i is higher than average in such states, $\beta > 1$ will make Y_i lower, and thus U_i' will be higher. Thus, the negative terms, although less in magnitude (taking into account the weighting by the p_i's) may still make the second term negative since it is possible that the they are weighted more heavily. The result does follow as before if it is assumed not merely that $\Delta\overline{B}' > 0$, but, more restrictively, that $\Delta B_i'' \geq 0$ for all i.[32] (Similarly, for the $\Delta\overline{B}' < 0$, the sufficient condition would be that $\Delta B_i' \leq 0$ for all i.)

Demonstration that $d=1 \Rightarrow b=0$: As noted previously, (*3) holds at this point and $b < 0$ has been ruled out. The previous demonstration ruling out $b > 0$ in this case relied on the demonstration that $\beta \leq 1$, which has just been shown problematic in this case.

Parallel proofs for $d > 1$: These follow the above pattern, the difficulty here concerning the inability to rule out $\beta < 1$.

2.3.3.2 Applicability of Proposition II

Lemma II.1 holds by the same logic as before, as does the argument ruling out $bK_b = \tilde{b}\tilde{K}_b$, the latter being apparent from the comparison of (*7) with (11) and (17). The only problem is that in ruling out $U_{kk}=0$, it was necessary to argue that the optimal b given $d=0$ was not greater than one -- Proposition I was relied on for that fact, but is problematic in this context unless B_o is defined in a manner such that B_o' represents either the highest or the lowest marginal benefit for any state for K in the relevant range. (Recall also that $\beta \leq 1$ was only a sufficient condition for $U_{kk} \neq 0$ to hold.)

2.4 Special Cases

The discussion thus far has examined cases where moral hazard is the central feature of the two-stage insurance model. This subsection considers two important contexts where risk-bearing considerations are unimportant,

----

[32] Alternatively, if the B_i are higher in the states where B_i' are higher, the problem would not arise.

which makes the simple externality point the entirety of the ex ante efficiency story.

2.4.1 Risk Neutrality and Diversification

This part of the discussion first considers the simple results in a risk-neutral world, and then discusses their applicability when the investor is an entity with widespread ownership.

2.4.1.1 Model When Investors are Risk Neutral

As one would expect, in the risk-neutral case, there is never any private insurance, or any risk-spreading benefit from government compensation. Therefore, the only effect of government relief is to distort incentives.

Proposition RN: *If $U''=0$, then $b^*=0$, sign $\partial U/\partial b$ = -sign b, $d^*=0$, sign $\partial U/\partial d$ = -sign d.*[33]

Proof: The first order condition for maximization over b is:

$$(8)\quad (1-p)\Delta B(U'_r - U'_o) - b\Delta B' K_b \bar{U}' = 0$$

$U''=0$ implies that the first term is zero. The sign of $\Delta B' K_b$ can be determined from:

----

[33] The derivative with respect to d is not really a partial derivative, but the notation is used to avoid the confusing alternative "dd".

$$(11)\quad K_b\Big[(1-p)U_o'B_o'' + (1-p)B_o'U_o''\cdot[B_o'-pb\Delta B'] + pU_r'[B_r''+(b+d)\Delta B''] + $$
$$+ p[B_r'+(b+d)\Delta B']U_r''\cdot[B_r'+(b+d-pb)\Delta B']\Big]$$
$$= p\Delta B(1-p)B_o'U_o'' - pU_r'\Delta B' - p\Delta B(1-p)[B_r'+(b+d)\Delta B']U_r''$$

The U″ terms on both sides drop. The remaining terms on the left side simply equal U_{kk}, and the only term on the right side is $-pU_r'\Delta B'$. Therefore, $\Delta B'K_b$ is positive.[34] From (8), the first two results follow.

For *d*, the first order condition can derived in the same manner as was the case for *b*, yielding:

$$(*1)\quad \frac{\partial U}{\partial d} = (\beta-d)\Delta B'K_\beta\beta_d\bar{U}' - \beta\Delta B'K_d\bar{U}'$$

Since b=0, regardless of *d*, β=d, and the first term equals zero.

For K_d, one can parallel the derivation of K_b, as follows:

$$(*2)\quad \frac{\partial^2 U}{\partial d\partial K} = (1-p)U_o'B_o''K_d + (1-p)B_o'U_o''\cdot[B_o'K_d-p\beta\Delta B'K_d-p\Delta B\beta_d] + pU_r'\cdot[B_r''K_d+\beta\Delta B''K_d+\Delta B'\beta_d]$$
$$+ p[B_r'+\beta\Delta B']U_r''\cdot[B_r'K_d+(1-p)\beta\Delta B'K_d+(1-p)\Delta B\beta_d] = 0$$

Paralleling (11), one can group to yield:

$$(*3)\quad K_d\Big[(1-p)U_o'B_o'' + (1-p)B_o'U_o''\cdot[B_o'-p\beta\Delta B'] + pU_r'[B_r''+\beta\Delta B''] +$$
$$+ p[B_r'+\beta\Delta B']U_r''\cdot[B_r'+(1-p)\beta\Delta B']\Big]$$
$$= \Big[p\Delta B(1-p)B_o'U_o'' - pU_r'\Delta B' - p\Delta B(1-p)[B_r'+(b+d)\Delta B']U_r''\Big]\beta_d$$

The left side bracketed term (following (23)) is simply:

----

[34] For (b+d) $\notin$ [0,1], it is possible that U_{kk}=0 at a maximum. Since U_{kk} can only approach zero from the negative side (it is only evaluated where *K* is maximized), K_b could be infinite, with the same sign as $\Delta B'$, which is sufficient for the argument to follow.

$$(*4)\quad \frac{\partial^2 U}{\partial K^2} - \beta(1-p)bB_o'\Delta B'U_o'\left[\frac{U_o''}{U_o'} - \frac{U_r''}{U_r'}\right]$$

Since $U''=0$ makes the second term zero and $\beta_d=1$, we have $K_d=K_\beta$. (Also recall that, at a given *d*, $K_\beta=K_b$.) Therefore, (*1) becomes:

$$(*5)\quad \frac{\partial U}{\partial d} = -d\Delta B'K_b\bar{U}'$$

This is the same as the first order condition for *b*, except that *d* substitutes for *b*. This establishes the last two results of the proposition. *Q.E.D.*

2.4.1.2 Dispersed Ownership versus Insurance

In many contexts of reform -- e.g., changes in the corporate income tax, OSHA regulations, and tariff policy -- the existence of dispersed ownership may be of far greater practical importance than insurance arrangements. For a widely-held public corporation, it may be appropriate, as a simplified first approximation, to think of the relevant decisionmaker as risk neutral, in which case the results just presented would be fully applicable.

There are a number of possible limitations concerning this assumption. First, systematic risk remains. But the government is not generally in a better position to spread such risk, as explored in subsection 4.2. Second, managers may be more risk averse than shareholders. In general, the information problems that contribute to moral hazard in the insurance context inhibit dispersed ownership as well since individual investors need some way to evaluate their investments and managers need incentives to behave in the investors' interests once the investment has been made. In fact, the concepts of insurance and dispersed ownership can be combined since the process of selling shares in a risky venture can be envisioned instead as involving two steps: the creation of an insurance company that bears almost all of the risk of the insured enterprise, followed by a sale of shares in that insurance company. Traditionally, we assume that insurance company managers perform the function of evaluating the risks of entities they insure in determining whether to offer a contract of insurance and how high to set the premium. This is analogous to the more general assumption that managers of firms behave in the interests of shareholders. If profit-sharing is used to align managers' and owners' interests, there will clearly be some problem with managerial risk aversion. Even if this were not the case, the processes by which managerial reputation is determined through ex post observation of results could produce similar behavior.[35] One might therefore observe firms purchasing insurance in the same manner as individuals, although Mayers and Smith (1982) have noted that there are many other possible explanations for such behavior that is observed.

Although this subject is rather complex and requires further attention, it seems a fair generalization that, in the context of widely-held corporations, considerations of risk will be relatively less important, making the ex ante incentive effects more likely to be decisive. Moreover, in examining transaction costs (see subsection 4.1), it should be noted that such considerations can be very important in the direct insurance context when there are many varied risks, each of low probability, whereas dispersed ownership deals with a wide range of risks simultaneously. Thus, relying on the market to handle government risks, when market risks are already sufficient to produce dispersed ownership in many contexts, may impose little additional transaction costs.[36]

----

[35] For this consideration, a different model for the evaluation of efficiency would be appropriate. The utility function used previously would be relevant for determining behavior (insurance and investment decisions), but a risk-neutral formulation would be appropriate for measuring welfare effects. The result is that government relief, which typically involves greater aggregate coverage than relying solely on private insurance, would be less desirable than suggested by the base case formulation because the additional risk-spreading benefits it provides would not be of social value, as was assumed previously.

[36] This discussion generally refers to "dispersed ownership" because it is focusing on the firm, rather than on diversification, which tends to focus on individual investors' portfolios (or the act of a single firm engaged in diverse enterprises). Of course, each is directly related to the other for the economy as a whole, in that one cannot have diversification if there is not dispersed ownership to begin with. Moreover, diversification is necessary for the argument in that individual shareholders would be significantly risk averse as to events affecting a particular firm if all their assets were

2.4.2 Insurance That Combats Moral Hazard

The basic model assumed that insurance policies involved the choice of a level of coverage (b), with the premium (a) being set on the assumption that the insurance coverage cannot directly be made a function of the level of investment (K) or the state of nature (s_i). I.e., it was assumed that only $\Delta B(K)$ could be observed. If the state could be observed directly, investors could contract for state-contingent payments that were independent of B_i, thus eliminating moral hazard. (This is simply the instance of there being a complete market for contingent commodities. [See Arrow (1971).]) Similarly, if the level of investment can be observed (either ex ante or ex post), transfers (in particular, *a*) could be made a function of K in a manner that eliminated the incentive problem. As a simple example, the premium for private insurance against a government taking would equal the probability of the taking multiplied by the amount of the investment; full coverage would not distort investment, as the risk of loss from additional investment would be fully reflected in the premium.[37] In either instance, private insurance, absent government mitigation, could achieve a first best.

It is rather straightforward to determine the effect of government compensation ($d > 0$) in either of these instances. In both cases, regardless of the degree of government relief, private insurance will cover the remainder of the risk, if any. But the "externality" problem will remain, leading to overinvestment in the simple case where $B_r' < B_0'$. The degree of distortion will vary directly with the level of compensation. Essentially, these special cases have properties much like the case where the investor is risk neutral. Government relief thus mitigates risk, but without preserving incentives,

----

invested in shares of that firm. In that event, even managers totally faithful to shareholders would behave in a significantly risk averse manner.

[37] Even if it is difficult to observe investment, so long as the insurance policy specifies a maximum coverage amount, the same effect is achieved. This is similar to the contingent commodity (states observable) case.

which can be accomplished by private insurance while spreading risk equally well.

2.4.2.1 States Observable

When states are observable, the result will be that insurance (in the absence of compensation) achieves a first best, because different premiums (lump sum payments, independent of ΔB, and therefore of K) can be collected (paid) in each state. To see this, simply allow premiums a_o and a_r in the two states, rather than constraining them to be equal. (2)(i) becomes

$$(*1)\quad (1-p)a_o + pa_r = pb\Delta B, \text{ or}$$

$$(*2)\quad a_o = \frac{p(-a_r + b\Delta B)}{1-p}$$

For convenience of notation, let $\alpha \equiv a_r$ and substitute for a_o using (*2) in applying the utility function.

<u>Proposition NMHS I</u>: *When states are observable, the insurance policy just described will produce a first best in the absence of compensation. For any* d, *b=0 and a_o and a_r determined by equations (*16) and (*17) are the unique optimum.*

<u>Proof</u>: The first order condition for K (4) is unchanged, since the premiums are taken as given (except that the evaluation of Y_i's will be different). The first order condition for b can be determined as follows:

$$(*3)\quad \frac{\partial U}{\partial b} = (1-p)U_o'[B_o'K_b - \frac{p}{1-p}b\Delta B'K_b - \frac{p}{1-p}\Delta B] + pU_r'[B_r'K_b + \beta\Delta B'K_b + \Delta B] = 0$$

The first U_o' term combined with the first two U_r' terms cancel, using (4). Regrouping, we have

$$(*4)\quad \Delta B(U_r' - U_o') - b\Delta B' K_b U_o' = 0$$

The first order condition for α is as follows:

$$(*5)\quad \frac{\partial U}{\partial \alpha} = (1-p)U_o'[B_o' K_\alpha - \frac{p}{1-p} b\Delta B' K_\alpha + \frac{p}{1-p}] + pU_r'[B_r' K_\alpha + \beta\Delta B' K_\alpha - 1] = 0$$

Similar manipulation to that used for (*3) yields:

$$(*6)\quad -(U_r' - U_o') - b\Delta B' K_\alpha U_o' = 0$$

Solving for $(U_r' - U_o')$ in (*6) and substituting into (*4) gives:

$$(*7)\quad -\Delta B(b\Delta B' K_\alpha U_o') - b\Delta B' K_b U_o' = 0, \text{ or}$$

$$(*8)\quad -(b\Delta B' U_o')(K_\alpha \Delta B + K_b) = 0$$

Therefore, either b=0 or $K_b = -K_\alpha \Delta B$. (Recall that $\Delta B'=0$ at an optimum directly implies no moral hazard, and was ruled out by assumption.)

To prove that b=0 is a unique optimum, this latter alternative must be ruled out. First, consider

$$(*9)\quad \frac{\partial^2 U}{\partial b \partial K} = (1-p)U_o' B_o'' K_b + (1-p)B_o' U_o''[B_o' K_b - \frac{p}{1-p} b\Delta B' K_b - \frac{p}{1-p}\Delta B]$$
$$+ pU_r'[B_r'' K_b + \beta\Delta B'' K_b + \Delta B']$$
$$+ p[B_r' + \beta\Delta B']U_r''[B_r' K_b + \beta\Delta B' K_b + \Delta B] = 0$$

Following the method used to derive (11) and (17), the terms can be grouped and substituted as follows:

$$(*10)\quad K_b[U_{kk} - pbB_o'\Delta B' U_o''] = pB_o'\Delta B U_o'' - pU_r'\Delta B' - p[B_r' + \beta\Delta B']\Delta B U_r''$$

Similarly,

$$(*11)\ \frac{\partial^2 U}{\partial\alpha\partial K} = (1-p)U_o'B_o''K_\alpha + (1-p)B_o'U_o''[B_o'K_\alpha - \frac{p}{1-p}b\Delta B'K_\alpha + \frac{p}{1-p}]$$

$$+ pU_r'[B_r''K_\alpha + \beta\Delta B''K_\alpha]$$

$$+ p[B_r' + \beta\Delta B']U_r''[B_r'K_\alpha + \beta\Delta B'K_\alpha - 1] = 0$$

Again, following the method used to derive (11) and (17), the terms can be grouped and substituted as follows:

$$(*12)\ K_\alpha[U_{kk} - pbB_o'\Delta B'U_o''] = -pB_o'U_o'' + p[B_r' + \beta\Delta B']U_r''$$

Let Φ represent the left side bracketed term in (*10) and (*12). Multiply both sides of (*12) by ΔB and add the respective sides to (*10):

$$(*13)\ (K_b + K_a\Delta B)\Phi = -pU_r'\Delta B'$$

If $b\neq 0$, the left side equals zero, which implies that $\Delta B'=0$ at the optimum, which, as noted before, was ruled out by assumption. Therefore $b=0$.

From (*4) (or *6), $b=0$ implies that $U_o'=U_r'$, which implies that $Y_o=Y_r$, i.e.,

$$(*14)\ B_o - a_o = B_r - a_r + d\Delta B$$

Also, from (*2), we have

$$(*15)\ a_o = \frac{-p}{1-p}a_r$$

Substituting (*15) into (*14) yields the expected result:

$$(*16)\ a_o = p(1-d)\Delta B$$

$$(*17)\ a_r = -(1-p)(1-d)\Delta B,\ \text{and}$$

In other words, the premiums spread all the loss that is not already spread through compensation. When there is no compensation ($d=0$), they spread all the loss, and when there is full compensation ($d=1$), there remains no loss to spread. *Q.E.D.*

Proposition NMHS II: *$d=0$ is the unique optimum. Moreover, for $d \in [-(1-p)/p, 1]$, sign $\partial U/\partial d$ = -sign d.*

Proof that $d^*=0$: Paralleling Lemma II.1, it is clear that the utility given any d is equivalent to that where $\tilde{d}=0$ and $\tilde{b}=d$. The insurance premiums will be such that $\tilde{a}_i=a_i+c$. Since the total of premiums for insurance and exaction for compensation are equal under both regimes, and β is equal, the same K will be selected. This in turn will produce the same utility. Moreover, by construction, $\tilde{a}_o$ is actuarially fair given the $\tilde{b}$ and a_r that have been selected. Since the utility is necessarily less than that at b=0, with premiums as described in Proposition NMHS I, as that proposition establishes, $d\neq0$ cannot be an optimum. *Q.E.D.*

Proof of sign $\partial U/\partial d$: Using the results of Proposition NMHS I, we can rewrite the utility function (3) as follows:

$$(*18)\ U = (1-p)U[B_o - p(1-d)\Delta B - pd\Delta B] + pU[B_r + (1-p)(1-d)\Delta B - pd\Delta B + d\Delta B]$$

Grouping the ΔB terms in both states yields

$$(*19)\ U = (1-p)U[B_o - p\Delta B] + pU[B_r + (1-p)\Delta B]$$

As indicated by Proposition NMHS I, income is spread evenly across states regardless of d (and thus d does not appear in this formulation). d will only affect utility through its effect on K, to be explored momentarily.

Begin with the first order condition for d:

$$(*20)\ \frac{\partial U}{\partial d} = (1-p)U'_o[B'_oK_d - p\Delta B'K_d] + pU'_r[B'_rK_d + (1-p)\Delta B'K_d] = 0$$

Since $U'_o=U'_r\equiv U'$, we have

$$(*21)\ \frac{\partial U}{\partial d} = [(1-p)U'B'_o + pU'B'_r]K_d = 0$$

In this case, (4) reduces to:

(*22) $\frac{\partial U}{\partial K} = (1-p)U'B_o' + pU'[B_r' + d\Delta B'] = 0$

Using (*22) to substitute in (*21) yields:

(*23) $\frac{\partial U}{\partial d} = -pd\Delta B'K_dU' = 0$

The simplest way to derive further information concerning K_d is to note that (*22) implies:

(*24) $(1-p)B_o' + p[B_r' + d\Delta B'] = 0$, or

(*25) $(1-p+pd)B_o' + p(1-d)B_r' = 0$

Take the derivative with respect to *d* and rearrange the terms:

(*26) $K_d[(1-p+pd)B_o'' + p(1-d)B_r''] = -p\Delta B'$

Note that, by assumption, the right side cannot equal zero, so $K_d \neq 0$. Moreover, $d \in [-(1-p)/p,1]$ implies that the left side bracketed term is negative, with the result that $\Delta B'K_d > 0$.[38] Examining (*23), this demonstrates the final part of the proposition. *Q.E.D.*

2.4.2.2 Investment Observable

When levels of investment are observable (either ex ante or ex post), the result will be that insurance (in the absence of compensation) achieves a

----

[38] It can readily be demonstrated that $d > 1$ is inferior to $d=1$. For any *d*, income is equal across states and total payments (compensation and insurance) are break-even. Therefore, utility can be compared simply be considering the real net benefits, $(1-p)B_o + pB_r$. This is simply a function of (K), with the optimum where $(1-p)B_o' + pB_r' = 0$. At d=1, $B_o'=0$, as noted on previous occasions (12). At $d > 1$, B_o' is negative (positive), and therefore *K* is greater if $\Delta B'$ is positive (negative). Since *K* is greater, and $\partial U/\partial K$ is negative in this range, it follows that utility must be lower.

A parallel argument can be applied to $d < -(1-p)/p$, because $d = -(1-p)/p$ implies that $B_r'=0$. The result will be that *K* is smaller in this range, but $\partial U/\partial K$ (defined as in the previous paragraph) is positive, again leading to lower utility.

first best, because premiums will take into account the level of K, eliminating the moral hazard problem. In particular, one can examine $a(K)=pb\Delta B(K)$. The first order condition for maximization of K no longer takes a as fixed, and instead substitutes $pb\Delta B(K)$ for a. Therefore, the problem reduces to the simpler one of maximizing utility choosing K and b, subject to the break-even constraint for a.

Proposition NMHK I: *When* K *is observable, the insurance policy just described will achieve the first best in the absence of compensation. And for any* d, *b=1-d.*

Proof: Begin with the new first order conditions:

$$(**1)\quad \frac{\partial U}{\partial K} = (1-p)U_o'[B_o' - pb\Delta B'] + pU_r'[B_r' - pb\Delta B' + \beta\Delta B'] = 0$$

$$(**2)\quad \frac{\partial U}{\partial b} = (1-p)U_o'[-p\Delta B] + pU_r'[-p\Delta B + \Delta B] = 0, \text{ or}$$

$$(**3)\quad \frac{\partial U}{\partial b} = p(1-p)\Delta B(U_r' - U_o') = 0$$

From this result it is clear that $U_r'=U_o'\equiv U'$, which implies that b=1-d. At d=0, this result eliminates risk and leaves incentives undistorted, as it is clear that (**1) reduces to

$$(**4)\quad \frac{\partial U}{\partial K} = (1-p)U'B_o' + pU'B_r' = 0$$

Q.E.D.

Proposition NMHK II: *The results of Proposition NMHS II hold.*

Proof: To prove that d=0 is the unique optimum, consider any d≠0. The same utility will be achieved at d=0, b=1 (which results in the same β) if K is selected at the same level as well. (Note that, given the definition of a, total income will also be the same.) This cannot, however, be the optimal K given d=0, as can be seen by reconsidering (**1) in light of the results of Proposition NMHK I:

(**5) $\frac{\partial U}{\partial K} = (1-p)U'B_o' + pU'[B_r' + d\Delta B'] = 0$

Since $\Delta B' \neq 0$ at the optimum (recall the assumption), the *K* that satisfies the equality in (**5) for any $d \neq 0$ will violate (**5) for d=0. The optimal *K* at d=0 will thus yield more utility than is possible for $d \neq 0$.[39]

For the sign of $\partial U/\partial d$, we can use the results of Proposition NMHK I to rewrite the utility function as follows:

(**6) $U = (1-p)U[B_o - p(1-d)\Delta B - pd\Delta B] + pU[B_r - p(1-d)\Delta B - pd\Delta B + \Delta B]$, or

(**7) $U = (1-p)U[B_o - p\Delta B] + pU[B_r + (1-p)\Delta B]$

Note that (**7) is the same as (*19) from the previous case. In addition, the first order condition for *K* (**5) is the same as the first order condition (*22) in the previous case. Therefore, the result for the sign of $\partial U/\partial d$ follows here as well. *Q.E.D.*

2.4.2.3 Discussion

There will surely be situations where these special assumptions hold, at least to a substantial degree. Since many reforms correspond to official announcements of legal change, it might be relatively easy to make coverage a function of the resulting state. The complication that frequently will arise, however, is that there are many varieties of possible reform, it being too expensive to specify all possibilities. Similarly, ex post, when the actual reform differs from those specified, there will be ambiguity concerning the amount of the transfer. Yet it might still be possible for private arrangements to reduce moral hazard through some degree of specification. In

----

[39] Alternatively, given the result to follow, one can apply footnote 38 (concluding Proposition NMHS II), which demonstrated that $d > 1$ resulted in less utility than at d=1, and $d < -(1-p)/p$ resulted in less utility than at $d = -(1-p)/p$. And within the interval $[-(1-p)/p, 1]$, d=0 is the maximum, completing this part of the proof.

some contexts, such as a complete taking (as for a highway) or in the event of a prospective change of great prominence,[40] there may be no difficulty. Another important possibility is that it may be feasible to index large numbers of states in a manner that is not affected by individuals' investment decisions. For example, a small defense contractor could make an arrangement for transfers contingent upon the overall level of the defense budget, or, if the contractor specialized, upon the amount budgeted in the future for a specific program. Or, for a Pigouvian tax, insurance payments could be made a function of the level of the tax imposed on the activity rather on the amount of the loss the insured realizes.

It also will be possible in some instances to observe K. In the example of a government taking, K may simply be the amount invested in improving land, which might be rather easy to document or appraise with a reasonable degree of precision. When the range of potential investments is varied and complex, and the effect of the government reform is to affect the value of prior investments in less straightforward ways, there will be limits to this form of contractual specification.

As a result, the base case, and those modifications to follow, rest on the set of assumptions most conducive to moral hazard problems, which may appear ironic since it is precisely such problems that make relief inefficient. In general, the greater the information the insurance company can observe -- even if the information is imperfect -- the less the moral hazard problem that will arise. [See Grossman & Hart (1983), Holmstrom (1979).] The base case model involved the extreme where both the investment level and states are not at all observable. The use of only two states was artificial in that the insurance company could have inferred the state of nature from whether or not a loss occurred. The assumption that it could not allows the simple, two-state model

----

[40] There is the reservation that the prominence of particular reforms of high probability may have evolved over time, which leaves the question of provision for the risk entailed by the earlier increases in the probability of change. As the introduction emphasized, it is changes in the probability of states -- including changes in the probability of change as well as ultimate realizations -- that constitute the relevant focus for analysis.

to capture the effects of the more cumbersome multiple-state formulation, which was briefly explored in subsection 2.3.3.

In these special cases where the insurance company observes either states or the level of investment, the difference is not that incentive effects do not exist, but rather that contractual specification can take them into account whereas mechanisms generally included in the description of transition relief cannot. Nothing in the analysis, however, suggests that government insurance, provided in a manner that mimics private insurance arrangements, cannot perform equally well. One argument for such government provision, discussed in subsection 4.4, arises when private actors misperceive the probability of adverse events, in which case compulsory insurance (which could be privately provided) might be desirable.

3. EXTENSIONS

3.1 Alternative Mechanisms for Transition Relief

In most contexts, one does not observe transition relief in the form of monetary compensation (or windfall taxation). More typical are mechanisms that directly nullify the reform to some degree. The practice of grandfathering involves exempting preexisting investment from a reform.[41] For example, a grandfather clause for the repeal of the income tax exemption for interest on municipal bonds might exempt the interest on bonds issued prior to enactment of the change. Grandfathering, like most transition relief, can often be partial -- in this context by granting only a partial exemption (e.g., from α percent of a new tax or subsidy) or by granting an exemption (whether it be partial or complete) for only a limited period of time. Grandfather clauses seem appealing because they avoid causing windfall losses -- in this instance to current holders of municipal bonds -- while at the same time they appear to avoid incentive problems because future decisions will be governed by the appropriate incentives.

Direct mitigation includes such mechanisms as partial implementation, delayed implementation, and phase-ins. Postponing the effective date of a reform serves to mitigate the risk imposed because some investments will no longer be affected after the passage of time (e.g., an expired bond)[42] and

----

[41] One common form of grandfathering has generally escaped notice. If, for example, a subsidy is given up front as a single payment, rather than as output is produced, then failure upon repeal to tax back the portion of the subsidy that corresponds to future output amounts to a grandfather provision. Therefore, initial program design can be an implicit part of future transition policy. For further discussion, see subsection 5.3.

[42] Of course, values can change in the interim due to the more certain expectation of a different regime in the near future.

others will be affected less in present value terms due to the delay. Partial (immediate) implementation is quite similar to delayed implementation in that it also has the general effect of scaling down both benefits and costs, including the incidence of risk.[43] (Phase-ins do not require separate analysis as they amount to combinations of partial and delayed implementation.)

The feature that distinguishes partial or complete grandfathering from partial or complete direct mitigation is that the former applies only to preexisting investment whereas the latter applies to *new* investment as well. This difference makes grandfathering preferable, in that it has the same effect on old investment without having to forgo the some of the benefits of the reform as to new investment -- those benefits being the reasons the reform was deemed appropriate in the first place. Of course, one could always convert direct mitigation schemes into partial grandfathering schemes. For example, a phase-in could be made applicable only to pre-enactment investment, with the new rules being immediately applicable to new investment.

Direct mitigation and grandfathering are alike as concerns *old* investment in two respects: First, the ex ante incentives created by each will be the same.[44] As a result, neither approach avoids the ex ante incentive problem that has been the focus throughout. In fact, these mechanisms are very little different from full compensation in this regard. Compensation covers the

----

[43] One additional relationship, noted by Zodrow (1981), is that immediate partial implementation will often be preferable to delayed full implementation (at a level providing the same total relief) because marginal net benefits will often be declining with the level of implementation (full implementation being defined as the point at which marginal net benefits equal zero). Delay, through discounting the entire future flow of costs and benefits, will tend to reduce net benefits proportionally; partial implementation will reduce benefits less than proportionately. Zodrow (1985) has noted the qualification that for some patterns of adjustment costs, the opposite preference can arise. Both effects concern investment after the announcement date, rather than ex ante considerations.

[44] This is not completely true to the extent that the relative scarcity created by a reform affects benefits. Since direct mitigation partially exempts new investment as well, direct mitigation of the same proportion as partial grandfathering would result in less of this scarcity effect. There will, however, generally exist a partial grandfathering scheme that provides the same aggregate relief as the direct mitigation scheme, taking the scarcity effects into account, which will be superior to the direct mitigation approach because of its more efficient treatment of new investment.

value the investment would have had if the relevant provisions had not been modified. Grandfathering amounts simply to exempting the investment from the modification, which has the same effect. Direct mitigation provides a partial exemption in this regard. Thus, for example, the promise that future reforms will contain comprehensive grandfathering provisions, like the promise of compensation, will distort incentives ex ante.

Second, both grandfathering and direct mitigation have the disadvantage, by comparison to immediate full implementation accompanied by equivalent compensation, that future benefits of the reform *as it applies to past investment* might be forgone. For example, it is better to compensate for the decreased market value created by a Pigouvian tax than to grandfather, because the latter approach -- although equivalent in terms of ex ante incentives -- destroys the corrective effect of the tax on future output. Similarly, if a product is sufficiently dangerous to be banned, a grandfather provision is clearly inferior to a ban combined with full compensation, and delayed implementation or a phase-in is clearly inferior to a ban combined with equivalent partial compensation.[45]

Therefore, with regard to future effects -- adjusting the levels of relief from the different approaches to hold ex ante effects constant -- there is a clear preference ranking among the alternatives: compensation is best, grandfathering is as good for new investment, but loses the future benefits that concern old investment, and direct mitigation is undesirable on both fronts with regard to future effects. Of course, the feasibility of grandfathering in any context depends upon the administrability of distinguishing between old and new investment, which in some contexts would no doubt be extremely problematic. In addition, as will be explored briefly in subsection 4.1, these alternative transition mechanisms are all generally

----

45 In cases where all relevant effects are in the past, grandfathering would be equivalent to compensation in terms of future effects. One example would involve repealing preferential depreciation deductions. For investment already made, being compensated for the loss resulting from reducing future deductions and grandfathering would have the same effect, since the future deductions are solely dependent upon past activity.

easier to administer than compensation because individual valuations need not be made.

The following discussion will continue in the spirit of the base case by confining attention to the ex ante effects on investment and risk. The basic approach will be to demonstrate the identity among different modes of transition relief. The first two subsections will consider simple forms of direct mitigation and grandfathering. The third subsection will briefly address complications that arise if it is assumed that the proportion of relief provided (i.e., the percentage of the loss or gain that is compensated or taxed) by a transition mechanism (regardless of type) varies with the level of investment, and is thus affected by private decisions. Finally, there is some discussion of the possible case for direct mitigation in situations where there exist multiple local optima in determining the appropriate level of implementation for the underlying policy.

3.1.1 Direct Mitigation

3.1.1.1 Partial Implementation

The simplest form of partial implementation would be one that involves only a given portion of the reform being adopted. For example, one may implement a proposed Pigouvian tax of only α times the originally proposed level.[46] In general, this would decrease ΔB. Although the degree of reduction need not be the same portion for any level of investment (K), it is instructive to examine the case where this strict proportion holds, as it is simplest to analyze completely. A priori, there is no reason to expect the

----

[46] Partial implementation must be defined by reference to full implementation. Here, full implementation means the reform that would be implemented ignoring transition considerations. In the simple case of the Pigouvian tax, it would be the level of tax (or subsidy) that would fully internalize the externality (and no more).

results to be qualitatively different if this strict relationship did not hold. (See subsection 3.1.3.) Define the benefits under proportional implementation as follows:

$$(*1)\ B_P \equiv (1-\alpha)B_o + \alpha B_r$$

As a result, income in the state involving reform will be as follows:

$$(*2)\ Y_r = B_P - a - c + b\Delta B$$

As before, $\Delta B \equiv B_o - B_r$. The "compensation premium" would be:

$$(*3)\ c = p(B_P - B_r)$$

Proposition DM I: *Partial implementation, as defined in (*1) - (*3), is equivalent to partial compensation, where d = 1-α, and thus is inefficient unless α=1 (full implementation).*

Proof: Begin with (*1):

$$(*1)\ B_P \equiv (1-\alpha)B_o + \alpha B_r$$

Add and subtract B_r from the right side:

$$(*4)\ B_P = (1-\alpha)B_o - (1-\alpha)B_r + B_r$$

Combine the terms:

$$(*5)\ B_P = B_r + (1-\alpha)\Delta B$$

Define $d \equiv 1-\alpha$, and substitute in (*2):

$$(*6)\ Y_r = B_r - a - c + (b+d)\Delta B$$

And (*3) becomes:

(*7) $c = pd\Delta B$

Thus, the problem is the same as in the base (compensation) case. *Q.E.D.*

3.1.1.2 Delayed Implementation: Simple Cases

Consider delaying the implementation of a "one-shot" reform -- i.e., one that involves a single action concerning the investment (such as a one-time levy, installation of a pollution control device, or condemnation). The effect on net benefits can be varied, depending on the time path of post-enactment date benefits. In the simplest case the effect of delay will be simply to defer an expenditure (or a subsidy) of a given amount, an amount which may well depend on the level of investment. Alternatively, delaying a reform with ongoing effects (such as an annual or per unit tax or subsidy) would have the same effect if the benefit stream resulting from an investment has the appropriate shape. For these simple cases, define the resulting benefits as follows:

(*8) $B_D \equiv B_o - e^{-\rho t}\Delta B$

where ρ is the discount rate and t is the length of the delay. Following the example of partial implementation,

(*9) $Y_r = B_D - a - c + b\Delta B$

(*10) $c = p(B_D - B_r)$

<u>Proposition DM II</u>: *Delayed implementation, as defined in (*8) - (*10), is equivalent to partial compensation, where* $d = 1-e^{-\rho t}$, *and thus is inefficient*

in regard to ex ante investment unless t=0 (no delay) or ρ=0 (delay has no effect on benefits).

Proof: Begin with (*8), and, as in Proposition DM I, add and subtract B_r, grouping as follows:

$$(*11)\quad B_D = (B_o - B_r)(1 - e^{-\rho t}) + B_r$$

Define $d \equiv 1-e^{-\rho t}$, yielding

$$(*12)\quad B_D = B_r + d\Delta B$$

As in Proposition DM I, (*6) and (*7) will result after substituting (*12) in (*9) and (*10), respectively. *Q.E.D.*

3.1.1.3 Delayed Implementation: A More General Formulation

A fully general formulation of delayed implementation would take into account that the benefit stream from an investment may have a variety of shapes, and thus the effect of the reform could vary depending upon the time period. One way to model this complexity is as follows:

$$(*13)\quad B_D(t) \equiv \begin{cases} B_o(t) & t \in [T, T_D] \\ B_r(t) & t \geq T_D \end{cases}$$

where T refers to the enactment date and T_D to the effective date.[47] It simplifies notation to define terms as follows:

$$(*14)\quad B_k^1(t) \equiv \int_T^{T_D} B_k(t)\,dt, \quad B_k^2(t) \equiv \int_{T_D}^{\infty} B_k(t)\,dt, \text{ for } k = o,r$$

B_D can now be rewritten as follows:

----

[47] This formulation is still not fully general, in that the streams $B_o(t)$ and $B_r(t)$ could be affected by T_D. For example, if there is substantial delay before implementation, it may be more efficient to close a plant at that time, whereas if delay were less, it may have been worth purchasing a pollution control device.

$$(*15)\ B_D = B_o^1 + B_r^2 = B_r^1 + B_r^2 + (B_o^1 - B_r^1) = B_r^1 + B_r^2 + 1 \cdot \Delta B^1$$

In other words, delay is equivalent to full compensation (d=1) for $t \in [T, T_D)$ and no compensation (d=0) for $t \geq T_D$.

This can be modeled with private insurance by assuming that it also can offer different levels of coverage for each time period. The utility function is

$$(*16)\ U = (1-p)U(B_o^1 + B_o^2 - a - c) + pU(B_r^1 + B_r^2 - a - c + \beta^1 \Delta B^1 + \beta^2 \Delta B^2)$$

where $\beta^i \equiv b^i + d^i$, the insurance premium is

$$(*17)\ a = p(b^1 \Delta B^1 + b^2 \Delta B^2)$$

and the "compensation premium" is

$$(*18)\ c = p(d^1 \Delta B^1 + d^2 \Delta B^2)$$

Because the problem (and, in fact, any mitigation scheme) can be modeled in a manner such that there exists an insurance scheme with no compensation (or mitigation) that results in the same level of investment and utility as that resulting from any compensation (or mitigation) scheme, with or without private insurance permitted, it again is clear that no improvement is possible. In the simple case, this was the result of Lemma II.1. Moreover, unless by chance the particular compensation scheme (taking into account the private insurance response) is mirrored by what would be the optimal insurance scheme in the absence of compensation, compensation (or mitigation) or any sort will be strictly worse. Such equivalence was proven impossible in the base case in Proposition II.

In this case, the results are even stronger, because the assumed increase in complexity generally permits private insurance to achieve a first best in the absence of compensation, and to result in full equalization of income. The reason is that the resulting scheme mimics the effects of insurance in the case where the states or the level of investment was observable. Essentially, because there are two insurance variables (b^1,b^2), they can be "played off" against each other, resulting in both the elimination of moral hazard *from insurance*[48] and full equalization of income. The idea is that the ratio b^1/b^2 can be set such that, at the optimal K, given (d^1,d^2), there is no moral hazard due to insurance. Then the absolute magnitudes of (b^1,b^2) can be scaled up or down as necessary to further equalize income.

Proposition DM III: *Delayed implementation in the presence of private insurance, as defined in (*13)-(*18), is inefficient unless, at the first best level of investment, the net effect of government relief on investment is zero -- i.e.,* $d^1\Delta B^{1\prime} + d^2\Delta B^{2\prime} = 0$. *Private insurance, defined by (*27), generally achieves a first best in the absence of any relief, and, regardless of the provision for relief, results in the equalization of income. The only exception is if, at the optimal K, given* (d^1,d^2) *and equalized income across states,* $\Delta B^1\Delta B^{2\prime} = \Delta B^2\Delta B^{1\prime}$. *In this case, private insurance does not equalize income, but no relief is still optimal, and generally is the unique optimum. It is necessarily the unique optimum when applied to the case defined in (*8) from the preceding subsection.*

Proof: The parallel derivations for this case begin with the first order conditions:

$$(*19)\ \frac{\partial U}{\partial K} = (1-p)U_o' \cdot [B_o^{1\prime} + B_o^{2\prime}] + pU_r' \cdot [B_r^{1\prime} + B_r^{2\prime} + \beta^1\Delta B^{1\prime} + \beta^2\Delta B^{2\prime}] = 0$$

----

[48] As in the earlier cases involving private insurance capable of eliminating moral hazard, the moral hazard from compensation is not eliminated because it is produced by an "externality."

$$(*20)\ \frac{\partial U}{\partial b^i} = (1-p)U_o'\cdot[B_o^{1\prime}K_{b^i} + B_o^{2\prime}K_{b^i} - pb^1\Delta B^{1\prime}K_{b^i} - p\Delta B^i - pb^2\Delta B^{2\prime}K_{b^i}]$$

$$+ pU_r'\cdot[B_r^{1\prime}K_{b^i} + B_r^{2\prime}K_{b^i} - pb^1\Delta B^{1\prime}K_{b^i} - p\Delta B^i - pb^2\Delta B^{2\prime}K_{b^i}$$

$$+ \beta^1\Delta B^{1\prime}K_{b^i} + \beta^2\Delta B^{2\prime}K_{b^i} + \Delta B^i] = 0$$

In (*20), the first two terms of each bracketed term, combined with the final two K_{bi} terms in the second bracketed term, can be dropped, as they add to (*19). The remaining terms can be combined as follows:

$$(*21)\ \frac{\partial U}{\partial b^i} = p(1-p)(U_r' - U_o')\Delta B^i - p[b^1\Delta B^{1\prime} + b^2\Delta B^{2\prime}]K_{b^i}\bar{U}' = 0$$

(*21) is equivalent to the first order condition for *b* in the base case (8), with the difference that the latter bracketed term here takes the place of $b\Delta B'$ in (8). The overall effect of the b^i terms is thus analogous to the effect of *b* in the simple case.

At this point, it can be demonstrated that the first order conditions are met if income is equalized and the moral hazard effect from insurance is zero -- i.e., $Y_r=Y_o$ and $b^1\Delta B^{1\prime} + b^2\Delta B^{2\prime} = 0$. (*21) clearly holds, so it is merely necessary to demonstrate that this is possible. (For the demonstration that this is a unique global optimum, see page 67.) Use (*19) to find the optimal *K* when those two conditions hold -- i.e., that solves:

$$(*22)\ (1-p)[B_o^{1\prime} + B_o^{2\prime}] + p[B_r^{1\prime} + B_r^{2\prime} + d^1\Delta B^{1\prime} + d^2\Delta B^{2\prime}] = 0$$

Now one can solve the two stated conditions for (b^1,b^2), at this level of *K*. From (*16), it is clear that equal income requires:

$$(*23)\ (1 - b^1 - d^1)\Delta B^1 + (1 - b^2 - d^2)\Delta B^2 = 0$$

The no moral hazard condition was simply:

$$(*24)\ b^1\Delta B^{1\prime} + b^2\Delta B^{2\prime} = 0$$

From (*24), the ratio of the b^i is simply

$$(*25)\quad \frac{b^1}{b^2} = -\frac{\Delta B^{2\prime}}{\Delta B^{1\prime}}$$

Substituting into (*23) for b^1 gives

$$(*26)\quad (1 + b^2 \frac{\Delta B^{2\prime}}{\Delta B^{1\prime}} - d^1)\Delta B^1 + (1 - b^2 - d^2)\Delta B^2 = 0, \text{ or}$$

$$(*27A)\quad b^2 = \frac{(1-d^1)\Delta B^1 + (1-d^2)\Delta B^2}{\Delta B^2 - \frac{\Delta B^{2\prime}}{\Delta B^{1\prime}}\Delta B^1}$$

Substituting in (*25) and solving for b^1 yields:

$$(*27B)\quad b^1 = \frac{(1-d^1)\Delta B^1 + (1-d^2)\Delta B^2}{\Delta B^1 - \frac{\Delta B^{1\prime}}{\Delta B^{2\prime}}\Delta B^2}$$

A number of comments are in order. First, the numerators in (*27) are simply that portion of the loss not already mitigated by government relief, which parallels the results in the cases involving no moral hazard because either the states or investment levels were assumed observable.

Second, the resulting moral hazard can be seen in (*22) to be simply that resulting from government relief. Thus, if d^1=d^2=0, a first best is achieved. Moreover, any government relief is inefficient unless $d^1\Delta B^{1\prime} + d^2\Delta B^{2\prime} = 0$ -- i.e., unless the effects of government relief in terms of incentives precisely cancel. In the standard case of delayed implementation -- in which d^1=1 and d^2=0 -- (or in more general cases involving phase-ins), this result will not hold unless $\Delta B^{1\prime}$=0 -- i.e., unless there is no moral hazard in the relevant time period to begin with.

A third and related point concerns the characterization of private insurance. From the denominators of (*27), or more directly from (*25), one can see that, in the typical case where the $\Delta B^{i\prime}$ have the same sign, b^1 and b^2 will have opposite signs. The b^i that is negative counteracts the moral hazard created by the one that is positive. Since observed transition schemes do not have this property, they do not take advantage of the benefit that can arise (i.e., in cases where transaction costs inhibit private insurance) from providing different levels of relief in different time periods.[49]

One added complication remains:[50] the solutions in (*27) required that the denominators not equal zero at the *K* that solves (*22). The condition for each is the same:[51]

(*28) $\Delta B^1 \Delta B^{2\prime} \neq \Delta B^2 \Delta B^{1\prime}$

From (*21), it is thus clear that optimal private investment will not equalize income across states. (If it did, the first term would be zero, and examination of the above derivation demonstrates that the second term cannot equal zero when (*28) holds, unless both K_{bi}=0 -- which is ruled out at page 69.) This case will now be examined in some depth, in part because (*28) holds in the simple cases involved in the previous subsection.

To determine the relationship of the b^i terms, the two first order conditions represented by (*21) can be combined. Begin by solving for $b^2\Delta B^{2\prime}$ in the first order condition for b^j:

----

[49] In practice, if ΔB^2 will be affected by subsequent investment during the first post-enactment time period (T,T_D), this result would be more complicated.

[50] There is also the complication involving U_{kk}=0 at the optimal *K*, which is discussed below in the case where (*28) fails. That analysis implies that the second term of (*21) could be infinite at the optimum just described. The second term is $-p[b^1\Delta B^{1\prime} + b^2\Delta B^{2\prime}]K_{bi}\bar{U}'$, which, in the case where U_{kk}=0, can be seen from (*33), below, to equal the right side of that expression divided by the inner bracketed term on the left side of that expression, and multiplied by $\bar{U}'$. When incomes are equal across states, that denominator will equal zero, and the numerator is the sum of $-pU_r'\Delta B^{i\prime}$ -- which cannot equal zero, except in the no moral hazard case -- and the second term, which does not necessarily equal the negative of the first. (Moreover, if the right sides equal zero for i=1,2, the derivation of (*34) demonstrates that (*28) would have to fail, ruling out the optimum in (*27), as described in the discussion to follow.)

This problem does not contradict the claim that (*27) -- if not ruled out by the failure of (*28) -- produces a global optimum. The analysis at page 67 demonstrates that (*27) results in the maximum possible expected income, given (d^1,d^2), as well as equal incomes across states. The preceding derivation indicates that no other combination of b^i's can have this result, which is obviously is the unique global optimum, if it can be achieved.

[51] If one (and only one) of the $\Delta B^{i\prime}$ equals zero, b^j is simply set to zero and b^i is adjusted to provide full equalization of income. (If both $\Delta B^{i\prime}$ equal zero, there is simply no moral hazard at the first best.)

$$(*29)\quad b^2\Delta B^{2\prime} = \frac{(1-p)(U_r' - U_o')\Delta B^j}{K_{b^j}\bar{U}'} - b^1\Delta B^{1\prime}$$

Substituting in the first order condition for b^i yields:

$$(*30)\quad \frac{\partial U}{\partial b^i} = p(1-p)(U_r' - U_o')\Delta B^i - p\left[\frac{(1-p)(U_r' - U_o')\Delta B^j}{K_{b^j}\bar{U}'}\right]K_{b^i}\bar{U}', \text{ or}$$

$$(*31)\quad \frac{\partial U}{\partial b^i} = p(1-p)(U_r' - U_o')\left[\Delta B^i - \Delta B^j\frac{K_{b^i}}{K_{b^j}}\right] = 0$$

If income is not equalized ($U_r' - U_o' \neq 0$),[52] it must be that $\Delta B^i K_{b^j} = \Delta B^j K_{b^i}$. An expression for K_{b^i} can be obtained from (*19):

$$(*32)\quad \frac{\partial^2 U}{\partial b^i \partial K} = (1-p)U_o'\cdot(B_o^{1\prime\prime} + B_o^{2\prime\prime})K_{b^i}$$
$$+ (1-p)(B_o^{1\prime}+B_o^{2\prime})U_o''\cdot[B_o^{1\prime}K_{b^i}+B_o^{2\prime}K_{b^i}-pb^1\Delta B^{1\prime}K_{b^i}-pb^2\Delta B^{2\prime}K_{b^i}-p\Delta B^i]$$
$$+ pU_r'[B_r^{1\prime\prime}K_{b^i}+B_r^{2\prime\prime}K_{b^i}+\beta^1\Delta B^{1\prime\prime}K_{b^i}+\beta^2\Delta B^{2\prime\prime}K_{b^i}+\Delta B^{i\prime}]$$
$$+ p(B_r^{1\prime}+B_r^{2\prime}+\beta^1\Delta B^{1\prime}+\beta^2\Delta B^{2\prime})U_r'' \times$$
$$\Big[B_r^{1\prime}K_{b^i}+B_r^{2\prime}K_{b^i}+\beta^1\Delta B^{1\prime}K_{b^i}+\beta^2\Delta B^{2\prime}K_{b^i}-pb^1\Delta B^{1\prime}K_{b^i}-pb^2\Delta B^{2\prime}K_{b^i}$$
$$- p\Delta B^i+\Delta B^i\Big] = 0$$

Substituting for U_{kk} and making the appropriate groupings, as in the base case, yields:

----

[52] If income is equalized, the implication is either that $K_{b^i}=0$, for i=1,2, or that $b^1\Delta B^{1\prime} + b^2\Delta B^{2\prime} = 0$. The second possibility has already been explored at length. The first possibility is ruled out in the discussion at page 69, equations (*37) - (*39).

$$(*33)\quad K_{b^i}\Big[U_{kk} - p(b^1\Delta B^{1\prime}+b^2\Delta B^{2\prime})[(1-p)(B_o^{1\prime}+B_o^{2\prime})U_o''+p(B_r^{1\prime}+B_r^{2\prime}+\beta^1\Delta B^{1\prime}+\beta^2\Delta B^{2\prime})U_r'']\Big]$$

$$= -pU_r'\Delta B^{i\prime} + p(1-p)\Delta B^i\Big[(B_o^{1\prime}+B_o^{2\prime})U_o'' - (B_r^{1\prime}+B_r^{2\prime}+\beta^1\Delta B^{1\prime}+\beta^2\Delta B^{2\prime})U_r''\Big]$$

The left side bracketed term is independent of i, as is the right side bracketed term. Also observe that the latter term is multiplied by ΔB^i. Thus, $\Delta B^i K_{b^j} = \Delta B^j K_{b^i}$ implies that

$$(*34)\quad \Delta B^{j\prime}\Delta B^i = \Delta B^{i\prime}\Delta B^j$$

The intuition behind this result is straightforward. Separate insurance coverage for the two time periods allows an investor to achieve a total level of risk spreading across the states through a combination of levels of coverage in each state. Each added "unit" of coverage for a given time period (each increment to b^i) increases insurance payments by ΔB^i, and adds to the moral hazard effect by $\Delta B^{i\prime}$. As a result, one would expect the optimum to be characterized by equality between $\Delta B^i/\Delta B^{i\prime}$ and $\Delta B^j/\Delta B^{j\prime}$, as (*34) suggests.[53]

It has now been demonstrated that the first order condition for the b^i (*21) implies either the local optimum described in (*27), or that described in (*34). If both are possible -- i.e., if (*34) fails at the K implied by (*22), but holds for some other combination of b^i and K -- it is clear that the optimum described in (*27) would be selected. The reason is that (*22) is the first order condition for the maximization of expected income, taking (d^1,d^2) as given, and (*27) equalizes that maximum expected income across states. As a result, the optimum implied by (*34) can have no greater expected income, and does not spread income equally across states, so the

----

[53] To further develop the intuition of this result, it is useful to note that substituting (*34) into the relationship for the moral hazard due to insurance produces the following equality:

$$(*N1)\quad b^1\Delta B^{1\prime} + b^2\Delta B^{2\prime} = \frac{\Delta B^{i\prime}}{\Delta B^i}[b^1\Delta B^1 + b^2\Delta B^2]$$

This holds for i=1,2, since (*34) implies that the ratios $\Delta B^{i\prime}/\Delta B^i$ are equal. The interpretation is that the moral hazard effect due to insurance (the left side) equals the moral hazard per unit of coverage (the ratio on the right side) multiplied by the total level of coverage (the bracketed term on the right side).

utility must be less than at the optimum produced by (*22) and (*27). Thus, the optimum described in (*27) is the global optimum for the b^i, so long as (*34) does not hold at the level of *K* implied by (*22), making it impossible.

Finally, we must consider whether $d^1=d^2=0$ (or, more generally, $d^1\Delta B^{1\prime} + d^2\Delta B^{2\prime} = 0$) is a *unique* global optimum in the special case where (*34) holds at the level of *K* implied by (*22). The argument of Lemma II.1 holds here following the same steps as in the base case. Thus, if there exists a global optimum (d^1,d^2,b^1,b^2), one can consider the alternative scheme where $\tilde{d}^1=\tilde{d}^2=0$ and $\tilde{b}^i\equiv d^i+b^i$, noting that the level of investment ($\tilde{K}$) and utility will be the same as that in the hypothesized global optimum. The first order conditions (*21) must hold in both cases, which implies:

$$(*35)\quad [b^1\Delta B^{1\prime} + b^2\Delta B^{2\prime}]K_{b^i} = [\tilde{b}^1\Delta B^{1\prime} + \tilde{b}^2\Delta B^{2\prime}]\tilde{K}_{b^i}$$

For convenience, designate the bracketed terms γ and $\tilde{\gamma}$ respectively. Following the derivation of Proposition II, we can construct the parallel to (18) from (*33):

$$(*36)\quad K_{b^i}[U_{kk} - \gamma\theta] = \Omega^i$$

The rest of the derivation of Proposition II from the base case can proceed from this equation, with the exception that γ plays the role of *b*. (The argument can be done for i=1 or i=2.) That derivation demonstrated that $b=\tilde{b}$, which was sufficient to imply that d=0. In this case, the analogous result is $\gamma=\tilde{\gamma}$, which does not immediately imply that $d^i=0$, for i=1,2. Subtracting γ from $\tilde{\gamma}$ indicates that $d^1\Delta B^{1\prime}+d^2\Delta B^{2\prime}=0$, the result advanced in the Proposition. Also, if $\gamma=\tilde{\gamma}$ in this case, one can use (*34) to substitute for $\Delta B^{2\prime}$, divide both sides by $\Delta B^{1\prime}$, multiply both sides by ΔB^1, and substitute for the $\tilde{b}^i$ to reach the conclusion that $d^1\Delta B^1+d^2\Delta B^2=0$. In other words, if the compensation in each period is precisely offsetting (at the optimal level of investment), so that net compensation is zero, both in aggregate and in terms of its marginal effect on the choice of *K*, one can still achieve a global optimum. This corresponds to the d=0 result from the base case.

The derivation of Proposition II needed to rule out $bK_b=0$ (here, corresponding to $\gamma K_{bi}=0$) and $U_{kk}=0$. First, consider the possibility that $\gamma K_{bi}=0$. From (*21), it is clear that this possibility implies equal incomes across states. If γ or $\tilde{\gamma}$ equals zero, one has the optimum described in (*27), where it has already been demonstrated that the Proposition follows. Here, consideration is limited to cases where (*34) holds, so this possibility is ruled out. The alternative is that $K_{bi}=0$, for i=1,2, which implies that $\Omega^i=0$. Examining the right side of (*33), the bracketed term in this case reduces to U'' times

$$(*37)\quad \Delta B^{1\prime}(1-\beta^1) + \Delta B^{2\prime}(1-\beta^2)$$

Equality of the Y_i's implies that

$$(*38)\quad (1-\beta^1)\Delta B^1 + (1-\beta^2)\Delta B^2 = 0$$

Using (*38) to solve for $1-\beta^2$ and substituting into (*37) yields:

$$(*39)\quad (1-\beta^1)\left[\Delta B^{1\prime} - \Delta B^{2\prime}\frac{\Delta B^1}{\Delta B^2}\right]$$

From (*34), the expression in brackets in (*39) equals zero, so, returning to the right side of (*33), it must be that $-pU_r'\Delta B^{i\prime}=0$, for i=1,2, which is impossible given the assumption that $\Delta B^{i\prime}\neq 0$ for some *i*.

Second, it was necessary to rule out the possibility that $U_{kk}=0$ at the optimal level of investment. As all the terms of U_{kk} are zero or strictly negative, except the third, a sufficient condition is that the third term be nonpositive:

$$(*40)\quad B_r^{1\prime\prime} + B_r^{2\prime\prime} + \beta^1\Delta B^{1\prime\prime} + \beta^2\Delta B^{2\prime\prime} \leq 0, \text{ or}$$

$$(*41)\quad \beta^1 B_o^{1\prime\prime} + \beta^2 B_o^{2\prime\prime} + (1-\beta^1)B_r^{1\prime\prime} + (1-\beta^2)B_r^{2\prime\prime} \leq 0$$

In the base case, the requirement was simply that $\beta B_o''+(1-\beta)B_r'' \leq 0$, which was implied by the result from Proposition I that, at any optimum where d=0,

$\beta \in [0,1]$. The analysis of Proposition I producing that result is not sufficient in this case to guarantee that $\beta^i \in [0,1]$, because it is possible for insurance in one (or both) time period(s) to be outside these bounds, offset by insurance in the other time period(s). For example, corresponding to the demonstration in Proposition I that, for d=0, b > 0, one can demonstrate in this case through the same reasoning that $b^1\Delta B^1 + b^2\Delta B^2 > 0$ (when the ΔB^i are positive). In other words, as in the simple case, the optimum insurance policy entails positive aggregate coverage. The difference here is that only aggregate effective coverage need be positive, whereas the sufficient condition for U_{kk} to be strictly negative (*41) required nonnegative coverage in each period. Similarly, the analysis used in Proposition I to rule out b > 1 in this case is sufficient to constrain the aggregate coverage, but not that for each period.

In the simple cases discussed in the previous section, however, the possibility that $U_{kk}=0$ can be ruled out. To demonstrate this result, the notation must first be translated. From (*8) it can be seen that $B^1_j=B_j$, for j=1,2, and $B^2_j=B_je^{-\rho t}$, for j=1,2 -- which implies that $\Delta B^1=\Delta B$ and $\Delta B^2=e^{-\rho t}\Delta B$. Loss uncompensated by the government is $(1-d^1)\Delta B^1 + (1-d^2)\Delta B^2 = 0\cdot\Delta B + 1\cdot e^{-\rho t}\Delta B = e^{\rho t}\Delta B = B_o - B_D$. Define $\lambda \equiv e^{\rho t}$. It is apparent that (*28) is violated and (*34) holds. Thus, the $U_{kk}=0$ problem arises. But analysis of the argument from Proposition I, applied to this case, is sufficient to bound the β^i to guarantee that U_{kk} is strictly negative. The demonstration that $b^1\Delta B^1 + b^2\Delta B^2 > 0$ implies that $\beta^1\Delta B + \beta^2\lambda\Delta B > 0$ when ΔB is positive, which in turn guarantees that $\beta^1 + \lambda\beta^2 > 0$.[54] Similarly, applying the reasoning used in Proposition I to rule out b > 1 in this context is sufficient to produce the result that $\beta^1 + \lambda\beta^2 < 1 + \lambda$. Finally, one can reinterpret (*41) by making the appropriate substitutions for this case:

----

[54] The β^i are used in place of the b^i, because for this demonstration (as it relates to proof of the Proposition) it is sufficient to consider the $\tilde{b}^i$ -- i.e., the b^i when d=0 -- in which case $\beta^i=b^i$.

$$(*42)\ (\beta^1 + \lambda\beta^2)B''_o + [(1+\lambda) - (\beta^1 + \lambda\beta^2)]B''_r \le 0$$

Given the restrictions just developed for the β^i in this special case, (*42) necessarily holds. The intuition behind this result is straightforward. In this special case, the problem reduces to the one period model, as one can see by defining $\Delta\tilde{B} \equiv (1+\lambda)\Delta B$ -- or, using the original notation of the simple case, $\Delta\tilde{B} \equiv (1+e^{-\rho t})\Delta B$. In the first order condition (*21), for the bracketed expression in the second term one simply has $(b^1+\lambda b^2)\Delta B'$, in which one could substitute $\tilde{b} \equiv b^1+\lambda b^2$, because everywhere the b^i appear in this case, it will be in this grouping. Given that the problem so reduces, the benefits possible through private insurance when (*28) holds are unavailable. The failure of (*28) is more general than this special case because (*28) is only applicable to the *K* determined by (*22), rather than throughout. In this special case, (*34) does in fact hold throughout, but, in general, this need not be true.

In the more general setting, there is no necessary implication that $U_{kk} \neq 0$, and there is no simple, intuitively appealing assumption that guarantees that result, so the analogy to Proposition II cannot strictly be advanced. Yet, the condition necessary for the corresponding proposition to fail has no a priori plausibility, giving no reason to affirmatively doubt the generality of that result. In addition, it should be noted that the problem of $U_{kk}=0$ is only relevant if the optimum described in (*27) cannot be achieved because (*28) fails at the *K* described in (*22). [This is demonstrated in note 50.] *Q.E.D.*

****

The standard case of delayed implementation is that where $d^1=1$ and $d^2=0$. It is clear that unless $\Delta B^{1\prime}=0$ at the optimum *K* -- i.e., that the reform would have had no effect during the period of the delay if there had been immediate full implementation -- the delay will be undesirable. The only possible exception would be where (*28) failed at the *K* described in (*22), making first best insurance impossible, and it also happened to be the case that, at the resulting global optimum, $U_{kk}=0$ (a possibility ruled out for the simple cases described in the previous section). If both these circumstances resulted, then delayed implementation would have no effect, as private insurance would adjust to produce the same net results as would occur through private insurance alone, in the absence of government relief.

3.1.2 Grandfathering

3.1.2.1 Simple Models Without Scarcity Effects

One form of partial (or complete) grandfathering would involve the immediate (and permanent) application of α of the reform to old investment. Ignoring scarcity effects, which are addressed below, benefits under grandfathering would be:

$$(*1) \quad B_G^P \equiv (1-\alpha)B_o + \alpha B_r$$

Alternatively, partial (or complete) grandfathering of the limited time period variety would be characterized as follows:

$$(*2) \quad B_G^D \equiv B_o - e^{-\rho t}\Delta B$$

The following results are immediately apparent.

Proposition GF I: *Partial grandfathering, through partial implementation with respect to old investment as defined in (*1), is equivalent to partial compensation, where $d = 1-\alpha$, and thus is inefficient with regard to ex ante investment unless $\alpha=1$ (no grandfathering).*

Proposition GF II: *Partial grandfathering, through delayed implementation with respect to old investment as defined in (*2), is equivalent to partial compensation, where $d = 1-e^{-\rho t}$, and thus is inefficient with regard to ex ante investment unless $t=0$ (no grandfathering) or $\rho=0$ (delay has no mitigating effect).*

Proof: (*1) is identical to (*1) in the direct mitigation section, so Proposition DM I applies. (*2) is identical to (*8) in the direct mitigation section, so Proposition DM II applies. *Q.E.D.*

These results follow directly from the conceptual identity between forms of partial (or complete) grandfathering and direct mitigation as they apply to old investment. The generalization of the delayed implementation formulation presented in Proposition DM III is similarly applicable.

3.1.2.2 Including Scarcity Effects

In many settings, grandfathering (partial or complete) has a greater effect than would result from not enacting (part or all of) the reform, because the reform is fully applicable to new investment. For example, if a reform increases the cost of an investment in a given activity, there will be less new investment than there would have been in the absence of the reform. Thus, over time, there will be an increasing gap between the investment that will exist given the reform and the investment level that would have existed in the absence of the reform, which would tend to increase the rental value on the old investment over time. As this effect is anticipated upon enactment, there will be an increase in the value of the old investment equal to the present discounted value of this increase in the rental stream. This affect will arise even in the absence of transition relief -- i.e., the unmitigated effect of a reform (B_r) will take into account the resulting change in investment patterns. With grandfathering, however, there is a greater scarcity effect to the extent that adverse impacts of the reform are inapplicable to old investment.[55] In the case of full grandfathering,

----

[55] As a simple, concrete example, the event of repealing the municipal bond interest exemption in the income tax code is often offered. In the event of repeal, the remaining supply of exempt interest bonds would decrease over time. Currently, the breakeven bracket for holding taxable and tax-exempt bonds (ignoring risk and related portfolio considerations) is well under the top marginal tax rate of 50%. The supply decrease will cause those holding such bonds in relatively lower brackets will sell to those in higher brackets who held bonds that have come due. As supply falls and the breakeven bracket

therefore, one would expect the change in net benefits upon enactment to be *positive* when the effect of the reform on new investment is *negative*, and conversely for the opposite case. In other words, full grandfathering provides more relief than does full compensation, and partial grandfathering would provide more relief than suggested by the formulations in (*1) and (*2).

The simplest way to model the scarcity effect (the same model can be used for both above cases) is to assume that the effects of (partial or complete) grandfathering are simply the sum of the direct effect, as described in the simple case, and a scarcity effect, which is assumed proportional both to the degree of direct relief (g) and the effect of the reform (ΔB).

(*3) $B_G \equiv B_r + g\Delta B + \sigma g\Delta B$

To correspond to the two simple cases, one simply has $g \equiv 1-\alpha$ and $g \equiv 1-e^{-\rho t}$ respectively. The scarcity effect is measured by σ.

Proposition GF III: *Partial grandfathering, as defined in (*3), is equivalent to partial compensation, where $d = g(1+\sigma)$, and thus is inefficient with regard to ex ante investment unless $g=0$ (no grandfathering).*

Proof: Defining $d \equiv g(1+\sigma)$ in (*3) yields

(*4) $B_G = B_r + d\Delta B$

This will produce the same value for Y_r as in the above cases, so those propositions apply. *Q.E.D.*

This scarcity effect arises because direct mitigation applies to new investment whereas analogous grandfathering does not. (This was also the source of the general superiority of grandfathering.) In general, this presents no significant problem for using grandfathering to provide a given level of relief, because the degree of grandfathering otherwise suggested in

----

rises, the effective interest rate will fall and the market value of such grandfathered bonds will rise.

formulations (*1) and (*2) can be scaled down by $1+\sigma$ to produce the same ex ante effect.[56]

The formulation in (*3) is obviously an oversimplification, in that it is not generally the case that the scarcity effect will have the simply linear form there modeled. This assumption is similar to the structure imposed on the models of the "direct" effects of grandfathering -- which were also assumed to have a strictly proportional effect. If either the scarcity effect or the direct effect of such relief varied (in terms of their per unit effect on ΔB) with ΔB, the propositions developed in the base case, which modeled compensation, would not be strictly applicable. That model would have to be generalized to permit d to be a function of ΔB. If a sufficiently general functional form were permitted, most results would be indeterminate.[57] Yet the effects described there would seem to hold in the contexts generally confronted. This issue will be considered further in subsection 3.1.3.

3.1.3 "Variable" Compensation

Thus far, both compensation and insurance have been assumed to take a simple form: each is modeled as paying (taxing) a stated percentage of the loss (gain), and breakeven is maintained by a state-independent transfer. The motivation for this modeling of compensation (and other transition relief) is

----

[56] To provide the same relief as suggested by α in (*1), simply define

(*N1) $\tilde{\alpha} \equiv 1 - \frac{1-\alpha}{1+\sigma}$

and to provide the same relief as suggested by t in (*2), simply define

(*N2) $\tilde{t} \equiv -\frac{1}{\rho}\ln\left[1 - \frac{1-e^{-\rho t}}{1+\sigma}\right]$

[57] The scheme discussed in note 58, interpreted as compensation -- i.e., $d \equiv \gamma\Delta B(K)^{-1}$. -- is instructive. Taking the derivative with respect to K and multiplying by $\Delta B'$ indicates that the moral hazard effect from compensation described by $d\Delta B'$ in (4) would be eliminated. Then, since γ is set equal to ΔB (evaluated at the first best K), a first best would be achieved. This is simply lump sum compensation.

in part that it offers a good approximation of how most typical transition schemes actually work. For insurance, the modeling may be less realistic, in that more complex schemes for insurance arrangements are often observed, and one can readily envision modifications that would be appropriate. For example, the premium could be partially state-dependent (for example, even if reform states cannot be fully distinguished, it may be possible to distinguish the set of states involving reform from those involving no reform) and the payment could be a stated function of the amount of the loss rather than a fixed proportion. In general, allowing more sophisticated insurance arrangements will permit them to better combat moral hazard, reinforcing the general conclusion that compensation will be undesirable.

Moreover, in the simple, two-state model that has generally been used, either modification would allow insurance to achieve a first best: the partially state-dependent premium scheme would amount to complete state-dependent payments and the variable compensation scheme would allow one implicitly to determine the level of investment. (For the latter, payments aside from premiums could be zero when the loss is zero and complete at the optimal level of loss, with the elasticity of payments with respect to the level of loss adjusted sufficiently to eliminate the moral hazard problem.[58]) These two special cases have already been described. The motivation for the framework that has been used in the rest of the discussion was to determine the effects of relief in instances where private insurance was imperfect.

In a model involving unobservable investment and more than two states, one could examine insurance or compensation schemes that were more complex in the manner described. The general results could change only if compensation were more sophisticated. The easiest way to illustrate this result is to consider the possibility that compensation is a function of the level of loss -- i.e., one now has $d(\Delta B(K))$, rather than d as a constant. If it is reasonable to

----

[58] If b could be any function of ΔB, the optimum would be $b = \gamma \Delta B^{-1}$, and zero at $\Delta B = 0$. Compensation would equal γ in the state involving reform, and γ could be set equal to $\Delta B(K^*)$, where K^* is the first best level of investment in the risk neutral case. This is equivalent to a lump sum, state dependent transfer, as would be provided if states could be observed.

make that adjustment, one might equally assume that insurance also had this additional property, and one would expect results much like those reached in the base case. The difference would be that, to the extent compensation better combatted moral hazard, the degree of inefficiency should be less.

To offer the simplest illustration of this effect, consider how the first order conditions change if one generalizes *d*, but not *b*.[59] (4) becomes:

$$(4')\ \frac{\partial U}{\partial K} = (1\text{-}p)U_o'B_o' + pU_r'\cdot[B_r' + (b+d+d'\Delta B)\Delta B'] = 0$$

which is identical to (5) except for the addition of $d'\Delta B$ in the last term. (5) becomes

$$(5')\ \frac{\partial U}{\partial b} = (1\text{-}p)U_o'\cdot[B_o'K_b - pb\Delta B'K_b - p\Delta B]$$
$$+ pU_r'\cdot[B_r'K_b - pb\Delta B'K_b - p\Delta B + (b+d+d'\Delta B)\Delta B'K_b + \Delta B] = 0$$

which is identical to (4) except for the addition of $d'\Delta B$ in the fourth U_r' term. But since that term drops when combined with other terms (as it is identical to the final term in (4')), the first order condition for *b* is not directly affected. (Of course, with a different *d* and *K*, the value of terms will be different.)

The most important effect of this modification on the base case derivations concerns the application of Lemma II.1, which established that there existed an insurance scheme that produced the same results, in terms of both investment and utility in each state, as any compensation scheme.[60] The proof of that lemma relied on the fact that the "equivalent" insurance scheme

----

[59] Keeping *b* as in the base case is done for simplification, and to focus on the effects of changing *d*. If *b* were a function of $\Delta B(K)$, then in place of the simple first order condition for *b*, one would be solving for the optimal function. As noted previously (see note 58), in this simple, two-period model, there would exist a solution that would be first best in the absence of compensation, which would leave one to analyze the affect of compensation alone, which is amply illustrated by the simple derivation to follow.

[60] The demonstrations in Proposition I concerning circumstances in which β is greater than, less than, or equal to one must be reassessed because of the added $d'\Delta B$ term in (4'), but these differences are of less interest. In addition, any characterization of *d*, such as $d > 0$, must be stated for a given value of *K*, such as that resulting when the first order conditions for *b* and *K* are satisfied.

provided not only the same level of compensation, but also still satisfied the first order condition for investment (4). In this case, if insurance remains of the simple form, there is no insurance scheme that is the complete analogue of a compensation. One could choose $\tilde{b} = b + d(\Delta B(K^*))$, where K^* is the level of investment that resulted at the optimal *b* given the function $d(\Delta B(K))$. The problem is that (4′) will not, in general, be satisfied at the same level of *K* unless $d'=0$, which was the assumption in the base case. To explore further, consider the case where $d > 0$ and $\Delta B > 0$.[61] If $d'>0$, one would expect the "corresponding" insurance scheme to yield *greater* utility than the compensation scheme, making the remainder of the proof of Proposition II unnecessary. Similarly, if $d'<0$, the corresponding scheme could yield *less* utility, which would disrupt the proof of Proposition II. The reason for these effects is that $d'>0$ exacerbates moral hazard, because increased investment not only increases the payment, given the level of coverage, but it also increases the level of coverage. Similarly, $d'<0$ combats moral hazard, because although increased investment still increases the payment for a given level of coverage, there is also a decrease in coverage that must be balanced against it.

Of course, if insurance were modeled analogously to compensation, there would be a fully equivalent scheme, and the analysis of Lemma II.1 would continue to hold. In any event, the implication for transition relief is clear. To the extent schemes systematically provide less relief (in proportion to the level of the loss) for greater losses, there will be less moral hazard -- and conversely for the opposite case.[62] As an overall policy, this may seem counterintuitive in that greater losses impose greater risk bearing costs. The idea here is that, for a given overall amount of coverage, it is desirable to have declining coverage at the margin. In other words, the marginal condition is a statement independent of the total amount of coverage that is provided. For example, a decreasing coverage scheme that had full

----

[61] For $\Delta B < 0$, the argument is reversed.

[62] Of course, if d' is sufficiently negative, there would be an investment distortion in the opposite direction.

coverage at the optimum level of investment might have more than complete coverage ($d > 1$) at lower levels of investment and partial coverage($d < 1$) at greater levels of investment.

As discussed in other sections, simple models of most forms of transition relief -- such as partial implementation, delayed implementation, phase-ins, and grandfathering -- have the property that $d'=0$. More generally, one does not strictly expect this to be the case. Delay may not necessarily have the same proportionate benefit at all levels of investment, and the scarcity effect described in connection with grandfathering need not be linear either. Yet, in all these cases, there is no apparent reason to doubt that these variations from $d'=0$ are typically, or even often, of a character to suggest that the results of the base case would be significantly misleading on the whole. In addition, if insurance is also assumed to be more complex, the overall distortion caused by transition relief could be worse.

3.1.4 Multiple Optima in Setting Underlying Policy

The argument that direct mitigation is dominated by partial compensation of a similar magnitude, and thus is undesirable, assumes that the policy instruments of the reform are continuously variable and that "full implementation" is the only point where marginal cost and marginal benefits (also assumed continuous) are equal. If all the stated assumptions are satisfied, market risk mitigation alone will be the optimal response. Consider the case involving the prospect of a future levy of a Pigouvian tax that will equal the external cost that is imposed by the private activity at that time. In order for the risk mitigation benefit to justify partial direct mitigation (i.e., the government implementing a tax less than the external cost), it must be that the resulting incentive cost is less than the benefit of further risk mitigation. But if moral hazard is the only factor preventing total risk mitigation by the market, it follows that any unmitigated risk imposes less of a cost than would result from the incentive distortion caused by greater protection. This reasoning restates the logic behind the argument that compensation (including partial compensation), to the extent greater than private provision, is undesirable and that direct mitigation will not be more desirable than partial compensation of a similar magnitude.

This argument can break down if the stated assumptions are violated. The reason is that if, for example, there are only two policy options (an all-or-nothing situation), it may be the case that total benefits exceed total costs by a small amount while, at the same time, the incentive effects from insurance coverage or dispersed ownership would be sufficiently adverse to produce significant risk-bearing costs. This can be the result because marginal costs may substantially exceed marginal benefits at the point of full implementation (scaling back to the point of equality is by assumption impossible). The same problem can arise if marginal costs exceed marginal benefits at a very small scale, making full implementation a nonunique local optimum, as it can also be the case that the benefits of full implementation exceed those of no implementation (or some substantially smaller level of implementation) by only a modest amount. In either case, even though full implementation is the global optimum in terms of conventional cost-benefit calculus that ignores the incidence of costs and benefits, it may no longer be the global optimum when risk considerations are taken into account. The earlier argument concerning incentives is disrupted because it may no longer the case that ΔB properly reflects the difference in social benefits attributable to the reform.

In the case of reform of major tax systems, variations in government demand, adjustment of substantive regulatory standards (such as permissible pollution levels or product content), or Pigouvian taxes and subsidies, it appears that the lumpiness problem would be absent and that the assumption that marginal benefits are nonincreasing throughout whereas marginal costs are nondecreasing would usually be a reasonable hypothesis. For a project such as building a highway, there is clearly some potential to vary the scope (e.g., the number of lanes), but it would no longer be the case that either assumption would strictly hold. Thus, there may be some instances where the optimal "transition" policy is complete abandonment, or a massive scaling down of the project.

3.2 Retroactive Application

Thus far, I have argued that incentive considerations favor a transition policy involving nominally prospective implementation with no compensation -- i.e., the new rules of the game should be immediately and fully applied to preexisting investments, but only as of the implementation date. In some instances, one might ask why the change should not be made retroactive. For example, if an activity is found to have negative external effects, the logic of the ex ante analysis suggests that it would be better that it be required to bear the corresponding cost for all its output, including that produced before the harm was discovered. To explore this case, it is necessary to explore further the meaning of "immediate" implementation.[63]

Until now, there has been virtually no mention of the effective date of a reform. The only exception was the discussion of delayed effective dates, but even there delay was discussed by reference to the enactment, with no attention being given to why that reference point was appropriate. The implicit assumption there, and in most discussions of transition relief, is that enactment is the relevant date, and, ideally, enactment would be instantaneous upon discovery that a different policy was appropriate. (Immediate enactment, of course, does not imply immediate application.) From this point of reference, delayed implementation appears as transition relief, and retroactive application (i.e., an effective date prior to enactment) would appear to be a transition penalty (in the case of a reform adverse to prior investment). This formulation, which begs the question of the appropriate time of implementation, is not always correct.

The easiest way to examine the issue is to focus on the ex ante incentive effects, as though one were in a risk neutral world. Consider two

----

[63] Compare the discussion of direct mitigation through partial implementation and that concerning multiple optima, both of which explored the meaning of "full" implementation.

justifications for reform: changed circumstances and new information about existing circumstances.[64] If circumstances change, making policy B preferable to existing policy A, it is desirable for policy B to take effect at the time the circumstances change. In cases where enactment follows the change almost immediately, making the enactment date the effective date is a reasonable approximation of this result.[65] Thus, for example, if a volcanic eruption has atmospheric effects that make future emissions of a substance harmful, when previously they were harmless, it would be appropriate to apply a newly enacted pollution tax to output subsequent to the change. Retroactivity would be inappropriate by assumption, since prior output was not harmful.[66]

Contrast the case of new information. Suppose it is discovered that an emission is harmful, and always has been. In this case, full retroactive application would be desirable. To take a simple case, consider an investor who, ex ante, believes that the emissions from the factory it is about to construct have a 10% chance of later being discovered to be significantly harmful. Without a policy of retroactivity in such circumstances, the investor will not expect to bear the social cost of its activities from their inception until the discovery date. By contrast, full retroactive application perfectly internalizes the externality.

Finally, consider the mixed case where new information changes government policies which themselves constitute new circumstances in the relevant sense. To make this rather abstract statement more concrete, consider the example

----

[64] Confining attention to these sources, rather than power or whim, is consistent with the general assumption of efficient government policy that is used in most of this investigation.

[65] Sometimes, the anticipation of modest retroactivity to align more precisely the effective date with the change may be important, as there otherwise might be a rush of undesired activity in the interim. For circumstances where the activity involves investment, with most of its effects spread over a long period in the future, when the new regime would be applicable, this will not be a significant concern.

[66] To the extent prior emissions remain in the atmosphere for a nontrivial period of time, it *would* be appropriate to have some retroactive application, covering emissions sufficiently recent to cause damage (and the amount of the tax, ideally, would be proportionately less for output further in the past). The analysis of new information, to follow, would be applicable to that extent.

where a local government determines that a new highway should now be built, and that it should have been built ten years ago. From an ex ante incentive perspective, it is inefficient to compensate for investments destroyed because they were in the path of a highway, because the social value of such projects is in fact zero, once they are leveled to allow the highway to be built. One might naively infer from the previous analysis of cases involving new information that in this case there should also be a retroactive tax on all the benefits the investment yielded over the past ten years. This is not, however, correct, as it can readily be seen that in the interim period any investment in the path of the future highway did in fact produce real social benefits, which, absent other imperfections, presumptively equal the private benefits that were realized.

In general, the test is rather simple. The "reference point" for determining when "delay" constitutes transition relief or "retroactivity" constitutes a transition penalty rather than part of the basic implementation of the reform is simply the standard test of equating private and social benefits to produce efficient private decisions. As a result, from this efficiency perspective, delay past the point of the change in the divergence of private and social benefits constitutes transition relief. In the simple changed circumstances case, an effective date after the change would be viewed as relief. In the pure new information case, even immediate implementation upon discovery of the implementation would constitute "delayed implementation" from this perspective, as an effective date retroactive to when the harm first existed would be necessary to fully effectuate the policy.

Given this analysis, no separate modeling of retroactivity is required. One can simply reinterpret the general formulation for delayed implementation presented in the discussion of direct mitigation. Recall equations (*13) - (*15):

$$(*13)\quad B_D(t) \equiv \begin{cases} B_o(t) & t \in [T, T_D] \\ B_r(t) & t \geq T_D \end{cases}$$

There, T referred to the enactment date and T_D to the effective date.

$$(*14)\quad B_k^1(t) \equiv \int_T^{T_D} B_k(t)\,dt, \quad B_k^2(t) \equiv \int_{T_D}^{\infty} B_k(t)\,dt, \text{ for } k = o,r$$

$$(*15)\quad B_D = B_o^1 + B_r^2 = B_r^1 + B_r^2 + (B_o^1 - B_r^1) = B_r^1 + B_r^2 + 1\cdot\Delta B^1$$

Delay was thus seen to be equivalent to full compensation (d=1) for $t \in [T,T_D)$ and no compensation (d=0) for $t \geq T_D$.

To generalize this formulation for the current discussion, simply redefine *T* to be the date when the real effects begin rather than the enactment date. Where there is simply a change in circumstances, and enactment is immediate, the interpretation is that offered previously, and $T_D > T$ is generally inefficient. In the case of new information, one might have T=0 (or T=-∞, depending on how time is indexed). In this case, implementation immediately upon discovery of the new information, but without a retroactive effective date, would produce the interpretation that T_D is the enactment date. For purposes of the efficiency analysis, however, all that matters is that $T_D > T$, so the inefficiency of failing to include retroactive application is precisely analogous to the use of delayed implementation is the simpler, changed circumstances case. For either situation, the net benefits in the base case model are simply $B_i \equiv B_i^1 + B_i^2$, for i=o,r.

A final word is in order concerning retroactive application. Consider again the above example where, ex ante, there was a 10% probability that the emissions would be harmful. Retroactive application results in a 10% chance of being charged for the full harm that might result. One might compare the imposition of an ex ante tax (i.e., one imposed with certainty) equal to 10% of the harm that might result. These alternatives are equivalent on an expected value basis, and thus are equivalent in a risk neutral world. In the presence of risk, however, the latter would appear preferable. In fact, if the investor purchased insurance against the 10% risk of the retroactive application, the premiums would precisely equal such an ex ante tax. The problem arises when there is a moral hazard effect due to the insurance option. The feasibility of the ex ante tax scheme is premised on the ability of the government to measure what is producing (probabilistically) the harm.[67]

As demonstrated previously, if this can be observed by the insurance company, premiums can similarly be gauged, with the result that moral hazard would be avoided. More generally, this government alternative raises the question of why such an approach to an uncertain future is not preferable to one of waiting to see what new circumstances and information are confronted, and acting at that time. The choice among these options will depend heavily upon the information available to the government ex ante (by comparison to that privately available) and administrative costs. This issue is discussed further in subsections 4.1 and 4.3.

Although one frequently hears that retroactivity is to be deplored, there are some aspects of government policy where retroactivity is the rule rather than the exception. Perhaps most notable are the common law rules applicable to tort and nuisance. If such common law rules change in response to new information, the general approach is to apply such rules to past behavior (subject to limits imposed by statutes of limitations, which themselves have often been relaxed to reach newly discovered harm). Alternatively, if the rule always has been that causing certain sorts of harm is actionable, it is often not a defense to the payment of damages for past harm that, at the time of the activity causing the harm, the harmful effects were not fully known. Some of the current disputes concerning liability for toxic substances and certain consumer products involve precisely these issues, in addition to more commonly aired complexities concerning causation and victim responsibility. This example is considered further in subsection 5.1.

3.3 Uncertain Transition Policy

There often exists uncertainty not only concerning the substantive content of future government policy but also with respect to the transition policy that would accompany a given future government policy, if enacted. Such uncertainty concerning transition policy can be viewed in at least two rather

----

[67] In fact, the feasibility of the ex post retroactive tax may depend upon the availability of similar information. Moreover, such information may have been available at the time the potentially harmful agent was produced or emitted, but difficult to extract long afterwards.

different ways. First, transition policy might be uncertain in the sense that the consistency assumption is violated. For example, a government may attempt to provide compensation or exact windfall taxes ex post that would not be optimal if anticipated ex ante. Such possibilities are generally outside the framework of this investigation and will be examined in subsection 4.6. Second, the government may not have yet decided what is the appropriate transition policy (viewed from the consistent, ex ante perspective). Or there may be ambiguity concerning the category in which a prospective reform falls (which would be relevant if different transition policies were used for different categories of reforms). Both circumstances might arise because it is costly ex ante to determine which transition policy is appropriate when there is a positive probability that the decision will never have to be implemented. There could also be mistakes or random elements affecting the selection. This latter group of possibilities are considered as a whole in this subsection. It should be noted, however, that to the extent a simple transition policy of no relief is, in the end, optimal in a wide range of settings, these latter possibilities may not be of great significance, whereas if other imperfections, such as those explored in section 4, were common, requiring a wide array of complex, context-specific approaches, the issue of uncertainty over transition policy in this second sense would indeed be crucial.

Intuition suggests that such uncertainty concerning transition policy is undesirable. Of course, in one sense, most uncertainty is undesirable. But in the realm of substantive government policy, some uncertainty is inevitable (and efficient), because available information and relevant circumstances change over time.[68] With respect to transition policy, however, such

----

[68] To a large extent, uncertainty is a question of cost. For example, in the case of a contemplated highway project, it may be costly to determine long in advance when a new highway will be necessary, or how it should be constructed. It should be noted, however, that even advance expenditures on information gathering and processing confined to the second decision can be quite valuable. For example, because of the benefits in channeling investment in the interim, it may be desirable to determine well in advance which route is to be selected contingent on there being sufficient need for a new highway, even if the analysis behind the selection is less accurate than may later be possible.

uncertainty can and generally should be avoided.[69] The point is simply that, since such uncertainty is unnecessary and avoidable, it should be eliminated.

As a simple illustration, assume that, with respect to a given set of reforms, there is a 50% likelihood that there will be full compensation and a 50% likelihood that there will be no relief. The problem in this case is that, in making the private insurance and investment decisions, the investor knows that in some states there will be less coverage than would have been desired and on others coverage will be excessive.[70] (Note that excessive coverage -- more than 100% -- is itself inefficient not only because additional moral hazard results, but because the effect on risk is adverse.)

This appears less desirable than two sorts of alternatives. First, it seems better (although obviously not first best) to divide equivalent reforms into two subsets, each with equivalent aggregate probability (assume they are mutually exclusive events) and announce in advance that one subset will receive full compensation and the other subset no compensation. Alternatively, one could simply make the uncertain decision in advance through a random process -- in the above example, one could flip a co un to determine whether there will be full or no compensation (with certainty). The ability to make different insurance and investment decisions based upon knowledge of which transition policy will be applicable increases efficiency.

Second, it seems better to receive the average level of relief with certainty. The reason is that, although income spreading between the reform and no reform states would be unaffected, there would be greater equalization of income across the reform states, increasing utility. Of course, one could

----

[69] To the extent information and circumstances which affect the choice of transition policy change over time, transition policy also would have to undergo changes to remain efficient. This subject is considered briefly in the section discussing the transition to an efficient transition policy. The point here is that there can be substantial uncertainty over efficient future substantive policy (such as the efficient level of Pigouvian taxes for various enterprises) without necessarily creating any uncertainty concerning the sort of transition policy that would be appropriate from an ex ante perspective.

[70] Even if private coverage could be conditioned on whether transition relief is provided, the result would be similar. In this example, coverage would never exceed 100%, but the two comparisons to follow would still be applicable.

simply note in either case that since the general conclusions thus far indicate that the optimal government policy entails no transition relief, these complications should not generally arise. The analysis of this section is meant to suggest, largely independent of those conclusions, that uncertain transition policy is generally undesirable, and probably dominated by clearly identifiable alternatives, regardless of what transition relief in fact is best.

As a simple application of this analysis, consider the possibility that transaction costs make insurance infeasible, suggesting that partial transition relief would be optimal.[71] Both alternatives outlined here would still be improvements over uncertain transition policy. The first alternative still permits the benefit of adjusting investment with the knowledge of which transition policy will be applicable. (I.e., the ability to self-insure is enhanced.) And the latter alternative still improves the distribution of risk across states involving reform. Of course, if the investor is risk-neutral, the only effect of relief is to create moral hazard, and the same adverse effect on incentives is produced in all the described scenarios.

This simple intuition is complicated by the fact that these alternatives not only facilitate risk spreading but also may affect the overall (or average) level of investment or insurance coverage, and these affects could be adverse. In general, the possibility that such affects are sufficiently negative to override the risk spreading benefits cannot be ruled out. Uncertain transition policy is clearly analogous to probabilistic insurance coverage, which is not what one observes. In addition, the issue presented here is also analogous to random taxation, where it has been demonstrated that it is possible for randomness to be desirable under certain conditions because it helps counteract adverse incentive effects that are analogous to moral hazard in this context. [See Hartman (1985), Newberry & Stiglitz (1982),

----

[71] In addition, to the extent that transition relief is motivated by the desire to avoid market transaction costs that are substantial, but not so great as to inhibit entering into insurance arrangements (see subsection 4.1), such a benefit might be lost entirely if there is not sufficient certainty ex ante concerning the availability of such relief.

Stiglitz (1982), Weiss (1976).] As a result, only limited formal analysis is offered. The remaining discussion covers the intuition behind the derivations, which are in Appendix B.

To be more specific on how these effects can arise, begin with the alternative of providing, with certainty, the expected value of anticipated compensation. This problem can be modeled as involving maximization of utility where the choice variable represents the proportion of compensation given in one of two states, each involving the reform being implemented -- where the proportions of compensation given in each state, weighted by the probability of that state conditional on reform, sum to one. (This model can be applied to any level of uncertain compensation, d -- i.e., it is not restricted to d=1, as in the above discussion.) If the stated proportion equals zero or one, one has the simple uncertainty case, and at some intermediate value one has the case involving equalization of income across "reform" states. To further simplify, and isolate on the effect of uncertainty in transition relief, continue to assume that private insurance in unavailable.[72]

As one varies the proportion from extreme values toward equalization of income across "reform" states, there are two primary effects. First, the greater equalization of incomes increases utility. Second, investment will increase. The latter result follows because changing the proportion shifts some of the moral hazard component (an increment of the $d\Delta B'$ term) from one "reform" state to the other. The state to which it is shifted is that with lower income and thus greater marginal utility of income. As a result, from the first order condition for maximization over K, one expects K to increase. In Appendix B, it is demonstrated that, if one is at an optimum proportion, it is indeed true (under base case assumptions) that investment increases as the proportion changes in the direction of greater equality across "reform"

----

[72] As noted previously, this indeed is an interesting case, since it is likely that partial compensation, rather than no relief, will be optimal. If insurance were also considered in the model, one would also have to take into account the adjustment in the levels of insurance coverage, in addition to adjustments in the level of investment. The spirit of the analysis would be much the same, although the results would be further complicated.

states. Since there is moral hazard present in this model, it is quite plausible that this increase in investment decreases expected income. The overall effect on utility is formally indeterminate. (In the appendix, it is demonstrated that equality is a local extreme point, but with straightforward assumptions it cannot be shown that it is a global optimum.) Alternatively, one could imagine providing equal coverage in the reform states, but at a different level than the expected value of d, in order to induce the same level of K as in the uncertainty case. Expected income would be the same, and this modification still improves risk spreading across states, but if the necessary level of d is lower, there is less risk spreading between the no reform state and the reform states, so the net effect on utility is indeterminate. As noted previously, however, since one does not observe contracts providing for probabilistic insurance coverage, there does not appear to be any direct evidence of the importance of this possibility that uncertainty is desirable.

Now consider briefly the other formulation -- where a lottery is conducted in advance to determine whether there will be compensation or where a subset of otherwise identical investments are designated to receive one level of relief and the remainder receive another level of relief. In this case, greater certainty in advance allows better self-insurance because investment can be better tailored to the level of coverage. Considering the case where projects are divided into two subsets, one can readily see the moral hazard issue. For the subset that is to get the greater level of relief, the incentive effect will be adverse, and for that which is to receive the lower level of relief, the incentive effect will be favorable. Since the greater and lesser levels are defined relative to the mean for purposes of comparison with uncertain transition relief, there is no a priori reason to expect the net incentive effect to be desirable or adverse. If, for example, the adverse incentive effect (measured in utility terms) were linear with the level of relief, then the hypothesized alternatives would be unambiguously preferable to uncertain transition relief. But no simple set of restrictions on the utility functions and benefit functions will guarantee this result. If the effect were adverse, one could imagine providing a lower level of relief (d) to induce the same K as in the pure uncertainty case. As with the previously discussed alternatives, the result is still favorable across reform states, but a lower d reduces risk spreading between the no reform state and the reform states, preserving the possibility that utility must decline. Once again, the result is formally indeterminate, although there is no particular reason to believe that there exists a readily definable set of instances where uncertainty in transition policy would be desirable.

3.4 Multiple Sources of Moral Hazard

3.4.1 Motivation and Intuition

The discussion thus far has suggested that, in the absence of market imperfections other than moral hazard, transition relief is inefficient. Until now, it has been assumed that there is only one source of moral hazard. It should be apparent that allowing multiple investment instruments would not itself change the results, even if the level of each investment is unobservable and insurance schemes are limited to covering a pre-determined percentage of the aggregate loss.

Consider the more complex scenario involving two sources of risk, one government and one "market."[73] Furthermore, add to the assumption that states are not observable ex post the related possibility that losses cannot be traced to their source. As a result, insurance covering b of the loss would cover the same portion regardless of whether it arose from government action or other sources. Government transition relief, however, is generally confined to losses directly resulting from government action. Such an assumption may appear problematic in the case of direct compensation, because making it dependent upon the source of the loss assumes that the source is, after all, observable. But other forms of transition relief, such as

----

[73] As elsewhere, "market" only means to indicate a risk not caused by uncertainty concerning future government action.

grandfathering and direct mitigation, typically provide relief without the need to measure and isolate the losses. Roughly speaking,[74] they provide relief by not taking (all of) the action that would cause the loss in the first instance. Thus, such transition relief offers an additional instrument. The discussion in this section considers the circumstances when such relief might be desirable in terms of ex ante incentives and risk, ignoring the previously discussed ex post inferiority of such alternatives in terms of future effects.[75]

To permit interpretations for a variety of possibilities, consider a model with two types of investment (K & L) and four states (s_{no}, s_{so}, s_{nr}, s_{sr}), where the first subscript indexes "nature's" action (*n* corresponds to "no action" and *s* corresponds to "adverse action") and the second indexes the government's action, as before. Furthermore, assume that *K* is observable by the insurance company and *L* is not. Finally, for the purposes of some of the discussion, it will be assumed that *K* is "directly" influenced only by nature's action (i.e., marginal net benefits are independent of the government's action) and that *L* is only thus influenced by the government's action.

Before proceeding with the derivations, consider intuitively the properties of this model. In the absence of government relief, the investor would purchase partial coverage: $b \in (0,1)$. If *L* were irrelevant (i.e., if its marginal net return were state-independent) or if it were observable, then the first best, with b=1, would be possible. If *K* were irrelevant, one would have the simple partial coverage solution of the base case. The level of coverage in this combined case would presumably reflect a compromise between these two simple cases.

----

[74] I.e., ignoring scarcity effects.

[75] Given the close connection between government and market risks in terms of justifications for government relief, it follows that if cases warranting transition relief can be justified along these lines, a similar case would hold for direct government mitigating action for parallel private risks. Corresponding to government mitigation of reforms, direct government mitigation of market actions might involve, for example, a government-imposed delay on the use of a new invention that caused losses to competitors, or protective quotas to protect existing investments against losses due to new foreign competition.

Consider the effect of adding government relief. In general, one expects less private coverage, and a net result that is inferior to the case of no relief. Here, government relief is inferior in part for the reasons suggested in the simple case where investment is observable. Since government relief is not contingent on *K* whereas the insurance premium (a) is, an element of moral hazard is introduced that otherwise would not be present. And even if *K* were not observable, government relief tends to induce excessive aggregate coverage.

In addition, as *b* is reduced in response to $d > 0$, less coverage will be provided for nature's action. The result is differential coverage for the two sources of risk. The remainder of this discussion is devoted to examining the effects of this. In the case where B_K (the derivative of *B* with respect to *K*) is only affected by nature's action and B_L only by the government, it would seem that government relief is undesirable. In fact, negative relief -- which would not always be feasible in this context[76] -- would be desirable.

The idea follows from the nature of the compromise in the no relief case. The investor is deterred from raising *b* all the way to b=1 because of the moral hazard effect with regard to *L*. Under the special assumptions stated, that effect arises solely due to government action. As a result, transition relief exacerbates the problem, in that it results in a situation where β -- the effective combination of *b* and *d*, $\beta \equiv d + b[1-d]$ -- is greater for losses due to the government's action that for losses due to nature's action. Negative relief would permit the opposite: ideally, it would allow for full compensation (b=1) for nature's action, which has no moral hazard effect, and optimal partial coverage -- $\beta \in (0,1)$ -- for the government's action.

It should be noted that a case favoring positive government relief cannot readily be generated by reversing the roles of *K* and *L*. Although making *L*,

----

[76] Negative grandfathering might be possible: for old investment, the rule could be changed even further than for new investment. For delayed implementation, one would have a negative delay -- i.e., making the reform retroactively applicable. To achieve the results of such a scheme would typically require a tax based on the effect of the reform, which, for purposes of this discussion has been assumed infeasible. Of course, such forms of "relief" are not observed in practice.

which is the unobservable investment instrument, dependent upon nature's act eliminates the potential benefits of negative relief just described, making *K*, which is observable, dependent upon the government's action does not generally make positive relief desirable. The reason is that the absence of a moral hazard problem with respect to *K* arises not because it is observable per se, but rather because the insurance premium can be made a function of *K*, creating an incentive to select it optimally. But to the extent coverage for losses affected by *K* are covered by government relief, like grandfathering or direct mitigation, which does not extract such an investment-contingent premium, it becomes more like the case where both *K* and *L* are unobservable. In that situation, there is no a priori benefit from the separation of insurance coverage between government and market risks in the direction indicated by positive relief, and there is the "externality" problem described in the base case that will affect total coverage, in addition to losing some of the benefits that arise because *K* is observable.

The case for positive relief along the lines described would arise where the losses caused by the government's action did not present a moral hazard problem in the first place -- i.e., where B_L, to return to the original scenario, is largely independent of the government's action. The pure case would be where the loss was totally independent of investment -- i.e., when it is "lump sum." The proposition that the government should compensate for losses that present no moral hazard problem is not subject to serious dispute on efficiency grounds, although such occasions may be rare. In this context, however, it seems a bit odd to hypothesize simultaneously that the losses from government's action are independent of private decisions and that such losses are not even approximately measurable independent from the effects of nature's action. Presumably, the assumed difficulty of measurement (prohibitively high cost) would typically arise because losses varied substantially with unobservable individual actions.

Although the argument just presented seems reasonably clear, formal analysis demonstrates that the results are somewhat less determinate. The reason is that there are numerous income effects which run in many directions, and not all of which vanish even if one assumes a constant absolute risk aversion utility function. As a result, the analysis to follow will not present any formal propositions, and the interpretation of effects will be limited to the most familiar case: where both government's action (r) and nature's action (s) are adverse in the sense that marginal and absolute benefits are lower (with the caveat that for some of the discussion, the marginal benefits will be assumed equal in the case of the investment instrument assumed independent of the action).

3.4.2 Analysis of the Model

The basic problem is a slight modification of that in the base case (1)-(2):

$\max_{d} U$, subject to:

(*1) breakeven with *c*

(*2) private maximization (optimal insurance decision):

$\max_{b,K} U|_{d}$, subject to:

(*i) breakeven with *a*

(*ii) $\max_{L} U|_{d,b,K}$

Income for each state is as follows:

$$
\begin{aligned}
& Y_{no} = B_{no} - a - c \\
(*3)\quad & Y_{so} = B_{so} - a - c + b(B_{no} - B_{so}) \\
& Y_{nr} = B_{nr} - a - c + (b+d-db)(B_{no} - B_{nr}) \\
& Y_{sr} = B_{sr} - a - c + d(1-b)(B_{so} - B_{sr}) + b(B_{no} - B_{sr})
\end{aligned}
$$

Note that the coverage is such that *d* only applies to the loss due to the government's action ($B_{io} - B_{ir}$). In addition, the fact that *b* is applicable to all observed loss does *not* simply result in a term equal to $b(B_{no} - B_{ij})$ for each state, s_{ij}, because it is assumed that the insurance company can only observe the total loss *that appears*, and since the government relief is assumed to be through grandfathering or direct mitigation, the total (unmitigated, uncompensated) loss must be reduced by that the government mitigates to determine the loss subject to insurance of *b*. Finally, the breakeven premiums are:

$$(*4)\quad c = d\sum_i p_{ir}(B_{io} - B_{ir}) = d\Delta\overline{B}^G$$

$$(*5)\quad b = b\sum_{ij}\sum p_{ij}(B_{no} - B_{ij}) - bd\sum_i p_{ir}(B_{io} - B_{ir}) = b\Delta\overline{B} - bd\Delta\overline{B}^G$$

(Note that $\Delta\overline{B}^G$ is the weighted average loss caused by the government's action.) The connection between these expressions for c & d and the explanation of the nature of the relief assumed to be provided in (*3) should be apparent.

As in the base case, there will be no direct attempt to derive $\partial U/\partial d$, because it would not be particularly illuminating. Instead, the approach will be a rough parallel to the logic of Lemma II.1 of the base case, which posited the alternative scheme $(\tilde{d},\tilde{b})$, where $\tilde{d}\equiv 0$ and $\tilde{b}\equiv d+b|_d$. That lemma demonstrated that the modified scheme produces the same level of investment and the same utility. The strategy of Proposition II was then to demonstrate that this modified scheme was not optimal given d=0, ruling out that the postulated *d* could be a global optimum.

Here, the transformation is more complex. Consider the following:

$$\tilde{d} = 0$$

$$(*6)\quad \tilde{b} = b + d(1-b)$$

$$\tilde{f} = -d(1-b)$$

The interpretations of $\tilde{b}$ and $\tilde{d}$ are as before. The interpretation of $\tilde{f}$ is that it is a private insurance policy (*K* observable) that provides coverage only in the event of nature's adverse state, and in the amount of $B_{no} - B_{so}$. Since the remainder of the discussion will involve these transformed schemes, the "~" will be omitted. As a result, the income levels in each state are:

$$Y_{no} = B_{no} - a - e$$

$$(*7)\quad Y_{so} = B_{so} - a - e + (b+f)(B_{no} - B_{so})$$

$$Y_{nr} = B_{nr} - a - e + b(B_{no} - B_{nr})$$

$$Y_{sr} = B_{sr} - a - e + f(B_{no} - B_{so}) + b(B_{no} - B_{sr})$$

And the breakeven premiums are:

$$(*8)\quad e = f \sum_j p_{sj}(B_{no} - B_{so}) = fp_s(B_{no} - B_{so})$$

$$(*9)\quad a = b\sum_i\sum_j p_{ij}(B_{no} - B_{ij}) = b\Delta\overline{B}$$

The problem is:

$$(*10)\quad \max_{b,f,K} U, \text{ subject to:}$$

(*i) breakeven with a,e

(*ii) $\max_{L} U$

<u>Proposition MMH I</u>: *The transformed scheme described in (*6) - (*9) produces the same utility and level of* L *if evaluated at the same level of* K *that was chosen in the original problem.*

<u>Proof</u>: Substituting (*6) into (*7) demonstrates that one has the same income in each state as in (*3), so long as the same *L* is chosen. But since the expressions for income in each state are the same, and the constraints result in the same total state-independent transfers (a+c in the original problem, and a+e in the transformed problem), the first order condition for *L* will be the same in each problem, and any value of *L* in one problem will produce the same utility as that value of *L* will in the other problem. *Q.E.D.*

Two additional comments are in order before proceeding further. First, in the transformed problem, it will generally be the case that a different *K* will be optimal than that produced in the initial problem. Since the transformed problem has no moral hazard in the choice of *K*, this adjustment will make the utility attainable from (*10) greater than that in the original problem, even before accounting for the fact that the *b* in the transformed problem would not generally be optimal either. Thus, the following will be true.

Proposition MMH II: *The transformed problem will generally produce greater utility than the original problem, so long as, originally, one does not have d=0 or the absence of moral hazard.*

Second, and more critically for the original motivation for this discussion, the insurance scheme (e,f) in the transformed problem was implicitly assumed to be infeasible in the opening discussion. The reason is that this adjunct scheme assumes that it is possible to measure $B_{no} - B_{so}$, whereas the original discussion posed the case where losses could not be traced to particular sources. But, if it could be demonstrated that f=0 was superior to f=-d(1-b), that would be sufficient to establish the superiority of no relief, since the transformed scheme would then involve only simple insurance.

Note that a negative value for *f* corresponds to a positive value for *d*. The intuition suggesting that $f < 0$ is undesirable parallels that given above: since the losses caused by nature only affect incentives with respect to *K*, which is not subject to moral hazard, a negative value for *f* only serves to eliminate desirable risk spreading with respect to those losses. It should be noted that instead of this transformed problem, one could alternatively examine a modified form of the original problem. One could maximize with respect to *d* assuming that $\beta \equiv d+(1-d)b$ remains fixed (i.e., assuming that *b* adjusts in this prescribed manner, rather than optimally, in accordance with the original problem). If, in this problem, it were shown that d=0 was preferable to any solution involving $d > 0$, the same result as discussed above with respect to f=0 would be applicable. In fact, it can be demonstrated that, aside from different treatments of *K*, this modification involving the maximization with respect to *d* is equivalent to the transformation presented above, when one maximizes with respect to *f*.[77] This transformation was chosen for further analysis because the notation is somewhat less cumbersome, and it provides some additional intuition into the nature of the problem.

The derivation begins with the first order conditions:

$$(*11)\ \frac{\partial U}{\partial L} = p_{no}U'_{no}[B_{no_L}] + p_{so}U'_{so}[B_{so_L} + (b+f)(B_{no_L} - B_{so_L})] + p_{nr}U'_{nr}[B_{nr_L} + b(B_{no_L} - B_{nr_L})] + p_{sr}U'_{sr}[B_{sr_L} + b(B_{no_L} - B_{sr_L}) + f(B_{no_L} - B_{so_L})] = 0$$

$$(*12)\ \frac{\partial U}{\partial f} = p_{no}U'_{no}[B_{no_L}L_f - a_L L_f - e_L L_f - \sum_j p_{sj}(B_{no} - B_{so})] + p_{so}U'_{so}[B_{so_L}L_f - a_L L_f - e_L L_f - \sum_j p_{sj}(B_{no} - B_{so}) + (b+f)(B_{no_L} - B_{so_L})L_f + (B_{no} - B_{so})] + p_{nr}U'_{nr}[B_{nr_L}L_f - a_L L_f - e_L L_f - \sum_j p_{sj}(B_{no} - B_{so}) + b(B_{no_L} - B_{nr_L})L_f] + p_{sr}U'_{sr}[B_{sr_L}L_f - a_L L_f - e_L L_f - \sum_j p_{sj}(B_{no} - B_{so}) + b(B_{no_L} - B_{sr_L})L_f + f(B_{no_L} - B_{so_L})L_f + (B_{no} - B_{so})] = 0$$

Using the first order condition for *L* (*11), a large portion of the terms cancel, leaving the following, simpler expression:

----

[77] If *K* is eliminated (or it is assumed that *K* is observable in setting *d*), the first order condition would be:

$$(*N1)\ \frac{\partial U}{\partial d} = \gamma(B_{no} - B_{so})\sum_j p_{sj}(U'_{sj} - \bar{U}') - L_d(a_L + g_L)\bar{U}'$$

Moreover, it is the case that $L_d = \gamma L_f$. Making that substitution, it can be seen that the first order condition for *d* in this modified problem is identical to that for *f* -- see (*13) below -- except for the factor of γ. In this modified problem, γ is simply the derivative of *f* with respect to *d*, which, from (*6), is clearly negative by construction.

$$(*13)\ \frac{\partial U}{\partial f} = (B_{no} - B_{so})\sum_j p_{sj}(U'_{sj} - \bar{U}') - L_f(a_L + e_L)\bar{U}'$$

The first term represents the benefit from further equalization of income across states differentiated by whether nature's action is adverse, weighted by the income differential subject to the adjunct policy. This term would be positive for $f \leq 0$ and $b \in [0,1)$. (If f becomes sufficiently positive, one will eventually reach the point where the net effect of further coverage is excessive in terms of spreading risk, with negative risk spreading consequences, analogous to the base case situation where $\beta > 1$.) The second term is the moral hazard effect. Note that

$$(*14)\ a_L + e_L = fp_s(B_{no_L} - B_{so_L}) + b\Delta\bar{B}_L$$

This expression is strictly negative, and in the special case simply reduces to $b\Delta\bar{B}_L$. Of course, at first glance, it might appear that L_f=0, if one examines the special case where nature's action has no effect on the marginal benefits from L. If this were true, the result would be that described intuitively at the outset, suggesting that $f < 0$, and thus $d > 0$, cannot be an optimum. It will be seen that, although the basic argument is true, there are additional effects to consider in determining the sign and magnitude of L_f.

To facilitate further investigation, it will be helpful to examine the first order condition for b as well:

$$\begin{aligned}(*15)\ \frac{\partial U}{\partial b} = {} & p_{no}U'_{no}[B_{no_L}L_b - a_L L_b - e_L L_b - \Delta\bar{B}] \\ & + p_{so}U'_{so}[B_{so_L}L_b - a_L L_b - e_L L_b - \Delta\bar{B} + (b+f)(B_{no_L} - B_{so_L})L_b + (B_{no} - B_{so})] \\ & + p_{nr}U'_{nr}[B_{nr_L}L_b - a_L L_b - e_L L_b - \Delta\bar{B} + b(B_{no_L} - B_{nr_L})L_b + (B_{no} - B_{nr})] \\ & + p_{sr}U'_{sr}[B_{sr_L}L_b - a_L L_b - e_L L_b - \Delta\bar{B} + b(B_{no_L} - B_{sr_L})L_b + f(B_{no_L} - B_{so_L})L_b \\ & \qquad + (B_{no} - B_{sr})] = 0\end{aligned}$$

Again, using the prior first order condition (*11) eliminates most of the terms, producing a simpler expression:

$$(*16)\ \frac{\partial U}{\partial b} = \sum_{ij}\sum p_{ij}U'_{ij}(\Delta B_{ij} - \Delta\bar{B}) - L_b(a_L + e_L)\bar{U}' = 0$$

The interpretation of (*16) parallels that of (*13). Here, the first term represents more straightforward insurance risk spreading. In the second term, one clearly expects L_b to be positive, given that the general moral hazard effect will be applicable with *b*, unlike with adjunct policy *f*.

It further illustrates the situation to solve (*16) for $(a_L + e_L)$ and substitute into (*13), to yield the following expression:

$$(*17)\ \frac{\partial U}{\partial f} = (B_{no} - B_{so})\sum_j p_{sj}(U'_{sj} - \bar{U}') - \frac{L_f}{L_e}\sum_{ij}\sum p_{ij}U'_{ij}(\Delta B_{ij} - \Delta\bar{B})$$

The double summation in the latter term will be positive, for *b* and *f* in the ranges described above, since the probability weighted sum of the ΔB_{ij} terms by definition equals $\Delta\bar{B}$, and the positive (and larger) terms correspond to greater losses (even taking mitigation and insurance into account), suggesting higher values for U'_{ij}.

The next step is to determine the signs of L_f and L_e. Since $\partial^2U/\partial f\partial L$ and $\partial^2U/\partial b\partial L$ have most of their terms group into U_{LL}, the direct determination of these derivatives will be skipped, moving directly, for L_b, to

$$(*18)\ U_{LL}L_b - (a_b+e_b)\sum_{ij}\sum p_{ij}\theta_{ij}U''_{ij} = -\sum_{ij}\sum p_{ij}U'_{ij}(B_{no_L} - B_{ij_L}) - \sum_{ij}\sum p_{ij}\theta_{ij}U''_{ij}(B_{no} - B_{ij})$$

where θ_{ij} is defined as the corresponding bracketed term in the first order condition for *L* (*11). For the first parenthetical, one has:

$$(*19)\ a_b + e_b = b\Delta\bar{B}_L L_b + \Delta\bar{B} + p_s f(B_{no_L} - B_{so_L})L_b$$

Substituting into (*18) yields:

$$(*20)\ L_b[U_{LL} - (b\Delta\bar{B}_L + p_s f(B_{no_L} - B_{so_L}))\sum_{ij}\sum p_{ij}\theta_{ij}U''_{ij}] =$$
$$- \sum_{ij}\sum p_{ij}U'_{ij}(B_{no_L} - B_{ij_L}) - \sum_{ij}\sum p_{ij}\theta_{ij}U''_{ij}(B_{no} - B_{ij}) + \sum_{ij}\sum p_{ij}\theta_{ij}U''_{ij}\Delta\bar{B}$$

Begin with the bracketed term on the left side. $U_{LL} \leq 0$ (and is strictly negative in the range under consideration). *b* and $\Delta\bar{B}_L$ are assumed positive. $B_{no_L} - B_{so_L}$ is zero in the special case where *L* is unaffected by nature's action.[78] Finally, using the assumption of nonincreasing absolute risk aversion, the double summation term is also nonnegative, following the argument used in discussing (21) - (23) in the base case. Therefore, the bracketed term is negative.

On the right side, the first term (including the leading minus sign) is strictly negative. The second two terms can be combined as follows:

$$(*21)\quad \sum_i\sum_j p_{ij}\theta_{ij}U''_{ij}(\Delta\bar{B} - \Delta B_{ij})$$

Without the term in parentheses, the expression is nonnegative. That term is smaller (and more likely to be negative) when the loss is largest, which corresponds to the case where the preceding terms (excluding the probability weight) is largest (and most likely to be positive). Since the probability weighted sum of the parenthetical term is zero, the expression is (*21) is negative. Therefore, the right side is unambiguously negative, so one has $L_e > 0$.

Following a similar procedure for L_f, the simplified expression for $\partial^2U/\partial f\partial L$ is:

$$(*22)\quad U_{LL}L_f - (a_f+e_f)\sum_i\sum_j p_{ij}\theta_{ij}U''_{ij} = -\sum_j p_{sj}U'_{sj}(B_{no_L} - B_{so_L}) - \sum_j p_{sj}\theta_{sj}U''_{sj}(B_{no} - B_{so})$$

For the first parenthetical term, one has:

$$(*23)\quad a_f + e_f = b\Delta\bar{B}_L L_f + p_s f(B_{no_L} - B_{so_L})L_f + p_s(B_{no} - B_{so})$$

Substituting into (*22) yields:

$$(*24)\quad L_f[U_{LL} - (b\Delta\bar{B}_L + p_s f(B_{no_L} - B_{so_L}))\sum_i\sum_j p_{ij}\theta_{ij}U''_{ij}] =$$

$$-\sum_j p_{sj}U'_{sj}(B_{no_L} - B_{so_L}) - \sum_j p_{sj}\theta_{sj}U''_{sj}(B_{no} - B_{so}) + \sum_i\sum_j p_{ij}\theta_{ij}U''_{ij}p_s(B_{no} - B_{so})$$

The bracketed term on the left side is identical to that in (*20), which was negative. On the right side, the first term would be negative if there were moral hazard with respect to *L* as to nature's action. This would tend to make

----

78 When the general case is discussed briefly, below, the possibility that this term is sufficiently large (recalling that $f < 0$ in this interpretation) to outweigh the $b\Delta\bar{B}_L$ effect, and in turn U_{LL}, will not be considered.

L_f positive as one would expect. In the special case, however, this term would equal zero.

Signing the combined effect of the two remaining terms is a bit more difficult, in general. Consider the special case of a constant absolute risk aversion utility function. From the first order condition for *L* (*11), it is clear that the third term would equal zero. The second term, which has the opposite sign of the second and fourth terms in (*11), will be negative (even without the additional restriction on risk aversion). The result is that, in general, L_f may well be positive. (If there is declining absolute risk aversion, the third term will be positive, and thus partly, or possibly completely, outweigh this effect, producing $L_f \leq 0$.)

From (*13) or (*17),[79] it is clear that there can be an incentive effect tending to make $\partial U/\partial f$ negative. (Inspection of these expressions, after making all obvious substitutions, yielded no further illumination concerning either necessary or sufficient conditions when $f < 0$.) The intuition can readily be derived from the components of the second term on the right side of (*24), which is easiest to understand by recognizing its correspondence to the second and fourth terms in the first order condition for *L* (*11). As adjunct coverage (f) increases, there is more income in the adverse states due to nature and less in nature's good states. As a result, the marginal utility of income increases in the good states and falls in the bad. From (*11), it is clear that the marginal returns are positive (or less negative) in the good states and negative (or less positive) in the bad states. The result is that the gain through a decline in risk imposed encourages the investor to increase *L*. (The lower level of *L* acts as "self-insurance".) But since *L*, unlike *K*, is assumed to be unobservable, the result is that *L* in fact must increase, leading to an increase in insurance premiums. Since moral hazard is present, the increase in *L* can lower expected income, and if there were such a reduction, and it was sufficiently large, it is possible in principle that the

----

[79] Note that if (*13) is used directly for making the interpretations, it is not necessary to assume that *b* is set optimally. In particular, like *K*, it could be assumed to be fixed at the level described in the transformation (*6).

benefit could outweigh the benefits of increased f in terms of risk spreading. And, if this net effect in turn decreased utility more than the increase achieved by allowing for the optimal adjustment of K and b in the transformed problem that involves no relief, it would be the case that relief could increase utility.

Note how this result appears possible even in the absence of a pure moral hazard effect (i.e., without $B_{no_L} - B_{so_L} > 0$). In assessing the plausibility of such a result, consider the following comparison. Suppose one begins with no government relief, and optimal private insurance of b. The government offers the investor a choice of the following: a state-contingent transfer (expected value equal to zero) that increases income in nature's adverse state or one that decreases income in nature's adverse state. For government relief to be desirable, the results here suggest that the latter alternative would have to be preferred. In general, however, one suspects that purely state-contingent transfers in the opposite direction (i.e., the direction that equalizes income across states) will increase utility. Of course, as demonstrated earlier for this model, and as is well-known, if a full state-contingent transfer were available, the investor could achieve the first best. Government relief in the case under discussion involves a large transfer in the opposite direction from that generally thought desirable.

Not to overstate this conclusion, it should be recalled that this particular, extreme characterization is fully applicable only if there is no moral hazard on L relating to nature's action. The first term on the right side of (*24) is positive when this assumption is relaxed, and measures the extent of that moral hazard effect. Therefore, the greater is the moral hazard with respect to L due to nature's action, the more plausible is the case for government relief. The extreme case, noted in the introductory discussion, is where all the moral hazard on L is due to nature's action. In this case, one would expect, a priori, that positive government relief might be desirable (one must account for the added moral hazard on K and the ex post inefficiency of government relief that is not pure compensation). This was the case, noted earlier, where is seemed more plausible that the loss attributable to government action would be capable of separate measurement, making this sort of modified model no longer applicable, returning one to the base case where even such relief would be unambiguously inefficient.

4. EVALUATION OF ASSUMPTIONS

4.1 Transaction and Administrative Costs

Insurance companies must charge more than actuarially fair premiums because they incur the costs of collecting information upon which to base premiums, writing policies, and determining the amount owed under the contract in the event of a loss. Additional, partially parallel costs must be borne by investors. Some such costs, at least for the insurance company, are spread over many policies, and investors may be able to rely in part on competing insurance companies for some of their information. As a result, in many instances such transaction costs will be sufficiently small as not to inhibit the sort of insurance coverage described thus far. Yet in other instances, particularly when there are very low probability events[80] or when the risks to be insured are smaller in magnitude (the latter being of less importance because the benefits from risk spreading are disproportionately less), one might expect such costs to inhibit coverage, or, alternatively, to substantially diminish the net benefits from insurance arrangements even if they still are made. Of course, diversified ownership spreads innumerable risks, generally without added transaction costs as more events must be covered.[81] None of these considerations in themselves suggest that the market will not achieve an efficient result in light of such costs.

----

[80] One of the most common responses to this problem is to group large numbers of events. Most familiar are life, health, and disability insurance which typically cover a virtually infinite array of causes, and umbrella liability policies. Dispersed ownership is the most prominent response in the business setting.

[81] Consideration of bankruptcy and some other costs makes this only approximately true.

Government insurance generally can do no better with these problems, but transition relief often can. The primary advantage of compensation over insurance in this context is that the former operates only ex post, thus incurring many such costs less often, or not at all. With low probability events that otherwise would have to be separately insured, this difference would be most important. It is, however, possible that the administrative costs of compensation awards are significantly higher. One reason, not necessarily inherent, is that valuations are typically processed through the legal system, whereas insurance payments often are not. An additional benefit of insurance in this respect is that the most important component of typical disputes -- valuation -- may have been determined in advance in setting levels of coverage. Finally, compensation must be funded. But the government already collects taxes, suggesting little administrative cost. Also, if compensation is joined with windfall taxation, positive net revenues may result.[82]

Compensation could be desirable in some instances if there were significant net administrative cost advantages and only modest costs in terms of ex ante incentives. If compensation were to be used in this manner, the earlier analysis suggests that it might be best to ban private insurance for such risks to avoid the externality that arises in the two-stage setting because of excessive aggregate coverage. This factor would only arise if the transaction costs were not so substantial as to preclude insurance.[83] In addition, the results of Proposition I make clear than compensation generally should cover less than one hundred percent of the loss.[84]

Finally, it should be noted that alternative transition mechanisms generally involve even less administrative costs than compensation. Not only

----

[82] See also the discussion in subsections 2.3.2 and 4.8.

[83] And, if compensation were complete, Proposition I indicates that no private insurance would be purchased in any event.

[84] Compensation for takings -- the one context where direct compensation is the standard transition rule -- purports to be full, but may be partial in practice, at least in some settings, because, for example, moving expenses and other adjustment costs are typically excluded, as is any idiosyncratic value of property, which may commonly exist in the case of houses.

are such mechanisms ex post, but they also do not require that valuations be made. There may be some additional costs. For example, grandfathering requires the ability to distinguish between old and new investment, which in some contexts might be extremely costly, and phase-ins will add compliance costs, as there will be more changes in the rules that actually govern. Yet one suspects that in many contexts (much of tax reform would be a notable example) such alternatives -- particularly delayed implementation or phase-ins -- would have virtually no administrative costs. Thus it appears that administrative cost considerations will tend to give precisely the opposite ranking among forms of relief than do ex ante incentive considerations. It should be highlighted, however, that when there is dispersed ownership, the ex ante incentive effects of relief are likely to be at their worst, as described in subsection 2.4.1.2, and transaction cost concerns are likely to be least important. Factors to be considered later in this section suggest that certain risks borne by individuals will be at the opposite end of the spectrum on both fronts.

4.2 Government as Superior Bearer of Risk

It is necessary to reconcile the conclusion of this section with Arrow and Lind's (1970) argument that public investment decisions should ignore considerations of risk (i.e., public projects should be evaluated using the discount rate for risk free private investments) although private investments do not. This result is troublesome for the conclusions here because it suggests that the government may be in a superior position to absorb risk than is the private sector. The explanation for this apparent inconsistency lies in the assumptions necessary for Arrow and Lind's result -- the one most relevant here being that the costs and benefits of the project be evenly spread over a large population. [See also Foldes & Rees (1977); Lind (1982).] Their illustrations assume that there is full exaction of benefits and full compensation for losses, and that the net benefits are then distributed evenly. But the entire transition policy issue arises precisely because there are often very unevenly distributed gains and losses from reforms, and if such distributional schemes create substantial incentive problems and administrative costs, it generally will not be the case that the theorem holds.[85] More generally, the question becomes why the government is more capable of distributing the effects in a risk-mitigating fashion than is the private sector. Private markets reach most of the population, and certainly most of those with any significant wealth, which is more relevant for risk-spreading purposes. It is unclear where the government would derive any significant comparative advantage in this context.[86]

4.3 Government Possessing Superior Information

One additional issue is that some sort of government intervention might be thought appropriate if the government believed it could assess the risks of future government action more accurately than could private institutions. Although at first glance it seems plausible that the government might be in a better position to assess the risks associated with its own future conduct than is the private sector, this conclusion seems doubtful. First, much of the risk concerning future government action arises from uncertainty concerning events in the world about which it is generally argued that private parties are in as good as or a better position to assess. A few examples illustrate this argument. Government farm policies may depend substantially

----

[85] Arrow and Lind do cite moral hazard and contracting costs as the two reasons that private markets do not offer complete insurance against risks, but do not indicate that these same problems would arise in attempting to conform government projects to the assumptions of their theorem. They do note, however, that to the extent risks associated with government actions are borne by private individuals, it would be appropriate to apply private discount rates, which include a risk premium. In addition, they suggest that in the case of large corporations (i.e., when the number of stockholders is sufficiently large), their theorem would also apply, and the corporation should behave in a risk-neutral manner, suggesting that they recognize that the superiority of government over private markets in allocating risk would be happenstance.

[86] Mayshar's (1977) argument that the government should subsidize risky projects to the extent it is less risk-averse than the private sector and shares in gains and losses through the tax system assumes some failure in the financial markets. He asserts that capital markets are in fact imperfect, giving the example of the lack of insurance for future wage income. To the extent moral hazard problems explain this phenomenon, however, his argument is deficient because government insurance faces the same problem. By contrast, if the primary difficulty were due to adverse selection, there might be a role for intervention.

on the weather or conditions affecting demand, where it might reasonably be assumed that the agriculture industry would be quite knowledgeable. Government energy policies depend upon factors that must be accounted for by the energy industry and consumers regardless of how the government acts. Government decisions to ban products depend upon the outcomes of testing that often are already known to the most affected private parties. Fiscal and monetary policy respond to a variety of conditions that are hardly unknown to major investors in financial markets.

Second, it is not obvious that government decisionmakers are in the best position even to assess what might be termed political risks -- e.g., who will win the next election, how likely the swing votes in Congress are to change before the close of the session, or the degree to which career bureaucrats will respond to a change in administration. Initially, much information is generally available to private parties, and the most affected parties make it their business to have the best available information concerning such developments.[87] In addition, many current policymakers probably make worse assessments than virtually anyone concerning the likelihood that their decisions will prove popular, result in reelection, or ultimately be deemed misguided or mistaken. It seems unlikely that in most cases the government would be in a sufficiently superior position in this regard than it is concerning general market risks to justify a substantially more activist policy in the context of governmentally-imposed risks.

More generally, the benefits of economies of scale and avoidance of disincentives due to external effects might place the government in a better position to assess risks, which would strengthen the argument for government intervention to mitigate market risks generally. There is no obvious reason that such benefits would be particularly applicable to risks concerning future government policymaking, and such information could be made available to private actors in any event. [Shavell (1984).]

----

[87] Sometimes they may have better information simply because they have off-the-record comments made to them alone or because they better understand their direct discussions and transactions concerning campaign contributions, bribes, and the like.

If any of these factors did suggest that, in a given context, the government possessed superior information, the optimal response would generally be government insurance -- perhaps compulsory insurance -- rather than transition relief. This will be explored further in the following subsection, which poses a particularly important instance involving superior government information.

4.4 Probability Misperceptions

All the formal analysis assumes rational, informed behavior. Yet it is well known that in many contexts, particularly those involving low probability events that involve large losses, individuals do not always behave rationally. [Kahneman, Slovic & Tversky (1982); Kunreuther et al. (1978); Kunreuther, Sanderson & Vetschera (1985); Nisbett & Ross (1980).] Perhaps the strongest case for government risk mitigation is presented by situations in which actors underestimate the probability of loss and thus do not purchase insurance when it is in their interest to do so. The risk-spreading benefits are obvious. Moreover, it is important to note that the ex ante incentive effects will be less significant. For example, if an investor mistakenly believes that the probability of an adverse event is zero, there will be overinvestment regardless of the coverage that is provided. More generally, if underestimation of probabilities were only partial, the incentive effects still would be less than in the base case.

Although this possibility substantially disrupts much of the argument, it is still not generally the case that transition relief is the appropriate response. Compulsory insurance would often be superior because of the requirement that a premium be paid. In those instances where premiums can be related in some way to investment levels, moral hazard can be partly or completely avoided. Even though the investor may still not understand or appreciate the probability of an adverse event, the fact that premium payments are tied to investment activity will be sufficient to induce efficient behavior.

This problem can be illustrated by considering the proposal that the government engage in a program of compulsory flood insurance. It is thought that many who build or buy in flood plains underestimate expected losses, and thus engage in excessive investment in such areas, while going without insurance coverage (even when subsidized insurance is available). [Kunreuther (1978).] If, for example, a condition for building in or moving into such an area was the advance payment of a year's flood insurance premium, even investors who thought disaster would never strike them would be induced to make efficient decisions. Moreover, in the event of a flood, there would no longer be a need for traditional "relief" because those suffering losses would already be covered.

This example involved a natural disaster, which has been grouped under the rubric of "market" rather than government risk in the earlier discussion. But, as noted in subsection 2.3.1, such risks are essentially the same for purposes of the issues considered here. The compulsory insurance seems unnecessary for many government risks, because the relevant actors will be businesses or sophisticated investors who are generally thought to be able to make adequate risk assessments. But when dealing, for example, with risks related to home ownership -- especially those that threaten large losses with low probability -- a compulsory insurance option appears to be desirable. Of course, because of the two-stage moral hazard problem, it might well be appropriate to ban supplemental coverage, particularly if a significant subset of the population is more sophisticated. Otherwise, the result for those investors would be that described in the base case, where excessive total coverage results.[88] Finally, if administrative costs made this remedy too costly, it seems possible that the second best policy might involve not only compensation, but some direct regulation -- e.g., limitations on development in the flood plain context.

----

[88] If the insurance is complete, as is efficient when the moral hazard problem can be controlled through premiums that are a function of the investment level, this consideration is irrelevant.

4.5 Adverse Selection

In sharp contrast to moral hazard, adverse selection problems can sometimes be directly corrected by government action, even when the government is not assumed to have superior information. The two conditions necessary for adverse selection to arise are (1) that the probability of loss (or, in some cases, the amount of the loss[89]) differs significantly among individuals, who are themselves aware of these differences, and (2) that insurance companies or other institutions be unable to detect those differences at a sufficiently low cost. [Akerlof (1970), Rothschild & Stiglitz (1976), Wilson (1977).]

Initially, it might appear that adverse selection is unimportant in the context of transitions, as most new policies are of broad applicability, ruling out the first requirement, and the probabilities of such events will not generally be the personal knowledge of particular affected actors, ruling out the second. In the presence of the kinds of information problems that give rise to moral hazard, however, there can be an adverse selection problem. For example, if one sought insurance against a possible product ban, the probability of the ban may be private knowledge both because of differential expertise and also because the investor's actions may determine product safety, thus affecting the probability of a product ban. If the relevant production functions are not sufficiently known by an insurance company, even typical partial coverage may not be possible, as insurance companies would be unable to estimate the expected losses from various ventures.

Government insurance offers a potential solution, as the government is not compelled to withdraw its protection when the lowest-risk groups drop out of the pool.[90] Transition relief offers similar ex post protection, without the premium. Not charging the premium will be generally be less important in

----

[89] If there were simply two groups, each having the same probability of loss, but with losses of different magnitude, there would be no adverse selection problem, as two policies offering different levels of coverage could be sustained in equilibrium. On the other hand, if the number of states is large so that, for example, each group has the same potential range of possible losses, but with different probabilities, the problem can arise. One could view this as involving different losses for the same probabilities, or different probabilities for the same losses -- the latter formulation making more clear how the adverse selection problem can arise.

contexts where adverse selection is a serious problem, because, by hypothesis, the information that might be used in adjusting the premiums is unavailable. In addition, there may be no two-stage moral hazard problem with supplementary coverage (if it is not banned) because, if the adverse selection problem is sufficiently great, it may be infeasible to sustain such coverage. It should be noted, however, that it is difficult to determine when government insurance or an equivalent approach is desirable. [Dahlby (1981).] In addition, since the adverse selection problem is most likely to arise in the context of government risks precisely when the moral hazard problem is serious, it seems even less likely that a government remedy will offer an overall improvement. Adverse selection in combination with moral hazard is yet another reason why dealing with risk through dispersed ownership rather than insurance may so often be desirable, and in the presence of dispersed ownership, as discussed previously, limitations in insurance markets will not be as important.

4.6 Implementing a Transition Policy

I have thus far ignored whether any particular government transition policy would be believed ex ante. It seems that conditions would often be conducive to the government's establishing a credible reputation since the situation is in essence an infinite time horizon repeated game. In addition, one can imagine ways, such as through the Constitution, that the government can make binding commitments.[91] The details and limitations of governmental institutions that facilitate the credible adoption of various transition policies will not be studied here, although a brief exploration of some of the factors relating to the credibility issue is useful.

----

[90] If one accepts alternative equilibrium concepts, such as those suggested by Grossman (1979) or Wilson (1977), a private insurance pooling equilibrium is more likely to exist, although it is still possible that a non-breakeven policy -- requiring government insurance or subsidization -- would improve welfare.

[91] The requirement of compensation in the event of a taking, as well as some other protections relating to government policy that affects property, is constitutionalized, although the potential for evolution through interpretation and the requirement of judicial enforcement places some limit on the constraint. [Kaplow (1986).]

In those instances where it is optimal to compensate (or otherwise mitigate risk) viewed from an ex ante perspective, it will necessarily also be optimal ex post. In theory, therefore, the government should not have a problem convincing private entities ex ante that compensation will be forthcoming.[92] Instead, the problem is that more compensation will often be optimal ex post than is appropriate ex ante. [Kydland & Prescott (1977).] So, contrary to what is often believed, the government's reputation problem consists of convincing parties that it will not in fact offer compensation for various risks that it imposes.[93] This view of the issue is in accord with examples such as preparation for the possibility of natural disasters, where it is commonly thought that the government will have little choice politically but to offer assistance after the fact.[94] Contrary views would thus have to be justified by various institutional considerations, which might reasonably be assumed to vary substantially case by case, depending upon the political influence of the affected parties. One suspects that historically many instances of compensation, grandfather provisions, and the like, can better be explained by political power and related issues that are considered in the following subsection, rather than by the issues that are the focus of this investigation.

Of course, in many contexts, the current situation in no way reflects the sort of transition policy argued here to be appropriate in the long run. This raises the issue of how to make the appropriate transition to a new transition policy. Since the argument for this approach rests on ex ante incentive effects, it obviously does not necessarily follow that immediate pursuit of a no mitigation policy would be appropriate. In light of the preceding point distinguishing the ex ante and ex post desirability of relief, it might appear

----

[92] This assumes, of course, that it is generally believed that the government's objective function demands that it behave optimally. This is considered in the next subsection.

[93] Similarly, ex post the government would be overly inclined to tax "windfall" gains. This relates to the argument that one-time expropriation of wealth would be optimal, so long as all were convinced that the action would not be repeated. (See subsection 4.8.)

[94] Of course, some sort of government policy, although not necessarily relief, can be justified, as discussed in subsection 4.4.

optimal to announce immediately that all future reform will be governed by such a transition policy, but provide compensation (and windfall taxation) for the reform entailed in the announcement itself. The adjustments would be substantial to the extent there are many reforms currently contemplated with both significant probabilities of enactment and prevailing expectations of transition relief. Moreover, it is not clear that such an announcement, superficially inconsistent as it might appear to some, would be credible. One could argue that credibility could only be established by announcement followed by immediate action denying relief in contexts where it was most expected. A more modest alternative ("delayed implementation" of the new transition policy) might involve, at a minimum, legislation or other pronouncements indicating that the new transition regime will be applicable in a given context as of a stated future date. Of course, in some areas (e.g., with tax reform), there are some established traditions concerning transitions, but, to a substantial extent, ad hoc compromises amidst political battles best characterize the current approach. Further analysis and speculation concerning this important issue will not be considered here.

4.7 Alternative Assumptions Concerning Government Behavior

All the basic results concerning the efficiency characteristics of ex ante incentive effects of transition relief were grounded on the assumed desirability of the underlying government policy. Since the effects themselves occur regardless of the basis for the underlying policy, it is clear that some assumptions concerning government behavior are necessary if any normative evaluation of transition policy is to be made. This subsection discusses a number of contexts in which alternative assumptions might be appropriate. In many instances, it is still the case that the ex ante efficiency properties will be much like those described before. To take a simple, extreme case, whatever the government's justification for leveling buildings for a highway, contingent on their being leveled, there are no further social benefits to be received from the destroyed investments. Even in such instances, however, transition relief might be beneficial if it has desirable effects on the underlying government policies that will be selected by the relevant decisionmaking body. The discussion begins with the most familiar context in which this issue has been raised, and then considers a number of other important settings.

4.7.1 Fiscal Illusion

The view that the government should not have to compensate for the costs it imposes may seem inconsistent with the general intuition deriving from the study of externalities that actors (here, the government) should bear the costs of their actions. This issue can be clarified by considering previous economic analyses of the legal system, including breach of contract [Shavell (1980a)] and tort law [Shavell(1979b, 1980b, 1982)]. The latter is the simplest to apply here. A longstanding dispute concerns whether liability for accidents should be covered by a rule of strict liability (the injurer always pays) or a rule of negligence (the injurer pays if and only if the injurer behaved negligently). In general, two conflicts arise. First, to the extent one party pays regardless of the behavior of the other, incentives for the other party are lost. Thus a rule of no liability gives insufficient incentives to the injurer and a rule of strict liability (with no defense of contributory negligence) gives insufficient incentives to the victim to avoid accidents. This aspect of the problem can be solved, but only by inquiring on a case-by-case basis to determine whether the injurer or victim behaved in an efficient manner, which can be both costly and subject to error. It is often sufficient to monitor the behavior of one party and -- so long as that party behaves efficiently -- have all remaining risk of loss borne by the other party. [See Shavell (1979, 1980b).]

If risk is to be borne by the government, the analogy for government transition policy would involve the government determining whether each private party eligible for compensation for losses invested at the appropriate margin (i.e., as if no compensation would be forthcoming) or an inappropriate margin (i.e., as if compensation would be paid). Such determinations would require that the government have all the information necessary to make what are normally private economic decisions, which I continue to assume is not feasible. But in the government compensation situation, this problem can be avoided. In the tort liability context, one (and only one) party's behavior must be independently scrutinized to determine whether it was efficient. But the government's decision concerning the appropriate reform has already been made, and is presumed to be efficient, so the instrument of compensation policy need only consider incentives of the private actors. Therefore, a policy of no compensation offers a simple solution that was not available in the accident setting.

But what if government reforms are not optimal? Requiring compensation might improve the incentives of the government agency. The "fiscal illusion" argument is that if the agency is not directly charged with the costs of actions it will discount them, and thus make inappropriate decisions. This commonly-noted point overlooks that the agency also typically does not directly receive the benefits, so the overall bias, if any, is unclear. If one thought most government action was undesirable, and that requiring compensation or some other transition relief would make all action more difficult, one might have little trouble in reaching a different outcome than that suggested here, although the same assumptions might also make one doubt whether there was much prospect for implementing such a transition policy.

Clearly more sophisticated models of government agency decisionmaking that highlight additional factors must be brought to bear to advance the argument. Two examples will be noted briefly. It is often suggested that the concentration of gains and losses will often affect the weighting they receive in the political process. [Downs (1957); Olson (1965).] This effect would suggest that, of average, there would be an excessive bias against projects requiring takings, because it will often be the losses that are the most concentrated.[95] Another view is that agencies seek to maximize their budget

----

[95] Some gains, as site-specific increases in land value, will be concentrated, but typically a more significant portion of the benefits would be dispersed relative to the costs.

or prestige. [Niskanen (1971).] In that case, requiring compensation might appear to increase an agency's incentives to undertake costly projects, as they would most increase its budget if they received the necessary appropriations. Requiring compensation, as well as imposing a windfall tax on benefits, might make the process more visible to a reviewing body, such as a legislative oversight committee, allowing for ex post verification of cost-benefit projections.

More generally, the impact of any particular transition policy on government decisionmaking is a complex question that is likely to vary substantially depending upon the context. In addition, the institutional question of government structure and internal incentives could not fully be evaluated without considering other, more direct approaches to these issues. Any consideration of this question, however, will no doubt have to take into account the ex ante incentive effects -- as suggested by the opening tort law analogy, where there is an incentive trade-off -- as well as the symmetry of losses and gains.

4.7.2 Political Feasibility

The desirability of a reform is often insufficient for it to be enacted. Losers may oppose based on self-interest, and have sufficient political power to prevent enactment.[96] Logrolling, which involves the combination of reforms that balance winners and losers, is one method by which opposition may be overcome. Another is transition relief, including direct compensation, to buy out the losers. If such relief is necessary to enact a reform, it may be desirable in spite of its adverse ex ante incentive effects. For example, if passage otherwise had zero probability, this would clearly be the case.

----

[96] Of course, nothing in the logic of this analysis suggests that gainers -- particularly modest gainers -- would not oppose enactment to the extent such opposition might result in a more favorable result, as would be the case if they were to receive added payments or special protections.

The analysis presented here suggests that providing such relief entails more than budgetary costs or the costs of forgoing benefits as to future actions. Moreover, there are broader ex ante effects of expecting transition relief in the event of political change: the incentives to influence government processes in the first instance might be increased. For example, if the general practice is to buy out losers from deregulation, the incentives to get favorable regulations (e.g., that largely redistributes wealth to a regulated industry, at some efficiency cost) enacted is increased because, even if such regulation proves short-lived, the promise of relief guarantees the future value of the private benefits in addition to any gained in the interim. [See Kitch (1977), McKenzie (1986).] Logrolling that combines desirable (or at least marginal) programs may be less of a problem because, so long as there is no transition relief for each reform, proper ex ante incentives will be preserved. This would be the case where similar groups (e.g., farmers, oil producers) benefited from one reform and lost from another, but each reform had different marginal effects on different particular investments.

This issue of political feasibility is another warranting further investigation. Understanding the efficiency costs involved in transition relief motivates the search for alternative methods of achieving political compromise, and for larger changes in government institutions that decrease the need to, in some manner, buy off opposition.

4.7.3 Changes in Government Direction Over Time

Another, partially related, relaxation of the government behavioral assumption employed here is to consider changes in political forces controlling government.[97] In this context, it might be thought that one way the government could mitigate risk attributable to future changes in policy

----

[97] One study considering this sort of problem is Schmalensee (1980).

would be simply to decrease the frequency with which such changes are considered. The degree to which this is the case varies depending upon the source of future changes in policy. Thus far, it has been assumed that changes in government policy reflect changes in existing circumstances or available information. To the extent this is true, the justification for moderating change is quite limited. Of course, the administrative and adjustment costs involved in changing policy as rapidly as possible to reflect every detail of the latest information would be substantial. But such considerations typically justify only modest delays, not major postponements of all substantial reforms, as some have advocated. These reasons and others for delay will be applicable at the time implementation is being considered -- i.e., from an ex post perspective -- and thus do not fall within the domain of transition policy issues being emphasized here.

It may be, however, that many changes in government policy have little to do with changes in available information, but rather have to do with characteristics of particular institutions. Consider the following hypothetical situation that is suggestive of a wider range of frequently encountered circumstances. There are two political parties, each with approximately 50% of the population in support, and each favoring radically different policies in a wide range of contexts due to their widely differing beliefs about the world. One could imagine either that their differences reflect different values (different objective functions) or different views concerning prevailing facts (i.e., they have the same social welfare function defined over characteristics, but different understandings of what would result in any given state).

Since the population is so evenly divided, and slight shifts occur over time, there is a substantial chance that each national election will bring a different party to power. How often should such elections be held? If each party could temporarily ignore who had the edge from currently being in power,[98] it might appear that they would agree to have rather infrequent

----

[98] Consider a Rawlsian "original position" perspective, or a constitutional amendment that will not take effect until a substantial time in the future,

elections. The more frequent the shifts in policy, the more difficult it will be for actions to be taken in reliance on any state of affairs continuing for any duration. As a result, more frequent elections will generally introduce biases toward short-term projects, even assuming that neither groups' beliefs exhibit a general preference for short-term rather than long-term investment. Therefore, ex ante, both parties might rationally prefer a system that lengthened the time between elections. On the other hand, very frequent elections, with frequently-resulting changes in power, might pose little uncertainty for very long-term projects because investors might be able to predict rather well what portion of the time each of various rules (and thus rewards) would govern, so no simple conclusions can be drawn.

It might be noted that compensation would be viewed in a different light in this sort of situation. One could construct a transition scheme with the result that investments made in any term would be little affected by which party was in power during the next. Thus, each term, the party in power would affect less decisions, but the effects of decisions made during their term would be longer lasting. The incentive problem would not be as applicable since the change in rules in later terms does not primarily reflect any actual change in the desirability of past policies. Of course, to the extent the party coming to power changes not merely incentive arrangements (e.g., taxes and subsidies) but direct actions (e.g., takings, government purchases), this qualification would not necessarily be in order. Compensation would result in greater investment ex ante, but since that investment would be destroyed, or produce output not used by the government sector (and valued less by the private sector since total demand is now less), is may be undesirable for precisely the reasons offered for the case where government reforms were presumed to be optimal.

Of course, in reality changes in policy, and in the party in power, reflect some combination of new information, changing values, and the luck of the political process. When widely differing positions are almost equally

----

where the probability of who will then be in power is nearly 50-50 and no term will be interrupted.

shared, slight changes, leading to modest numbers of converts, or chance events might be sufficient to produce substantial reforms. In general, it would appear that some conservatism -- in the sense of decreasing the frequency with which major changes in the status quo were considered -- might be wise.[99] On the other hand, as noted previously, at the time such political shifts occur -- even in the more extreme situations illustrated in the hypothetical example -- it would be known that the reforms are likely to be temporary, and thus the effects would be much less. It is unclear whether this countervailing consideration is in practice sufficient to dislodge completely the suggestion that the frequency of change should be moderated to some degree.

A more direct response would be better. At least to the extent different policy positions reflect different information, it seems preferable to develop decisionmaking processes that reflect some compromise among positions (e.g., in proportion to the fraction holding each view) rather than using all-or-nothing majority decisionmaking rules since the latter result in greater change (and thus greater imposition of risk) not warranted by the magnitude of change in probability estimates suggested by slight shifts in support for various positions. (This argument is independent of the benefit that might come from taking advantage of the information held by all.) As discussed in subsection 3.1.4, a large portion of government policies are, roughly, continuously variable. Such policies could be adjusted modestly, the reform being in proportion to the degree of change in underlying beliefs. The basic idea is that there are benefits to government processes that more precisely reflect the extent of majorities (as well as the strength of beliefs). To some extent, ordinary legislative compromise, as discussed in subsection 4.7.2, accomplishes this result. That discussion differed because there it was taken for granted which reforms were desirable, and the question was how

----

[99] One could argue that slight shifts in support typically reflect only slight shifts in information, and thus warrant only slight adjustments in what may be a closely balanced cost-benefit comparison. Allowing policy to respond frequently to such slight changes in information thus provides only limited benefits whereas, if the risks imposed are significant and not diversifiable, it may impose significant risk-bearing costs.

to achieve support from losers. Here, the issue directly concerns which policies are desirable.[100] Finally, if one believes that the political process generates results largely in accord with a median voter model, frequent changes in the governing party would not generally be associated with large policy shifts, so the problems discussed in this subsection would not arise.

4.7.4 Abuse of Power and the Need to Constrain Government

Another potential justification for transition relief in some contexts is that it may help to restrain pure abuses of power. For example, the Constitution prohibits ex post facto legislation in the criminal context, as well as bills of attainder -- i.e., punitive legislation that targets particular individuals -- presumably based on the fear that too many such instances of government action will be ill-motivated. The requirement of compensation for takings could be viewed similarly: although it does not prohibit the action which targets particular individuals, it requires compensation, which minimizes the harm that can thus be inflicted (and, in turn, substantially reduces the incentive to use takings in a punitive manner). Similarly, general constitutional requirements of equal protection inhibit ill-motivated action against powerless minorities. The majority may inflict whatever harm or deny whatever benefit it chooses, so long as it treats itself in the same manner. Finally, as discussed at the outset, complete transition relief nullifies government action designed to raise revenue or determine the distribution of income or wealth. Thus, strong requirements of relief effectively limit government in both these respects.

----

[100] There is an important overlap to the extent pure self-interest defines desirability, in that large losers will have values and views that are opposed to change whereas winners will favor it. Of course, real shifts in the number of gainers and losers (holding average gains and losses constant) does affect the desirability of the reform. When disagreement is attributable in significant part to different beliefs, the issue is rather different.

The general difficulty in using transition policy to achieve the end of preventing abuse is that it is very difficult in the abstract to define with sufficient precision the categories of action that are most subject to abuse without also foreclosing desirable or essential government policies at the same time. A broad prohibition on punitive redistribution, for example, might effectively foreclose all potential to raise revenue. The ideal solution would be an independent body that could determine which decisions were properly motivated and which were not. Of course, if such a body existed, the problem would vanish, as it could be given the final authority for approval of government reforms, thus ensuring that they were appropriate.

The equal protection provision in constitutional law has raised precisely this difficulty, the current result being, roughly, that unequal treatment that differentially affects particular groups (e.g., classifications based on race, religion, and, to some degree, sex) are carefully scrutinized and frequently prohibited. In contrast, virtually all other bases for differential treatment are permitted, precisely on the ground that the courts do not wish to be in the business of second-guessing legislative and executive motives when there are conceivable legitimate bases for the actions (which the courts do not feel sufficiently expert to scrutinize). [Kaplow (1986).] These questions have occupied political scientists and legal scholars for ages (and economists with increasing frequency), and no doubt will never be fully resolved.

4.8 Relation to Confiscatory Taxation and Raising Revenue

It has frequently been noted that "ex post" taxes are desirable because they raise revenue without accompanying distortions. The extreme example is the idea that the government should suddenly take the entire capital stock, and announce that it will never do such a thing again. (Given all the revenue, it also may be able to do without all future taxation of capital income as well.) The problem is said to be the credibility of the latter statement in light of the former action, which relates directly to the issue explored in subsection 4.6 concerning the credibility and implementation of transition policy.

Unlike the case of taxation for revenue-raising, it is generally desirable that government policies be anticipated, and induce ex ante behavior that takes them into account, as suggested by the entire discussion thus far. Transition relief, of course, would prevent ex ante distortions attributable to the prospect of new revenue-raising initiatives, but it would have this effect precisely to the extent it nullified such change. Transition considerations for an increase in a Pigouvian tax and an increase in the income tax (to raise revenue or affect distribution) are thus opposite: ex ante anticipation and ex post compensation (i.e., if ex ante effects could be ignored) are both desirable for the former and undesirable for the latter.[101] The reasons, however, are parallel. Both ex ante considerations concern incentive distortions and both ex post considerations concern distribution.[102] These issues will be considered further in subsection 5.4.2's discussion of tax reform.

To the extent it could be determined that certain sorts of revenue-raising would never be desirable, it might be beneficial for the government to commit itself not to rely on them in the future if there is a non-zero probability estimate concerning such future action. Taxes are not, however, unique in this regard. It is always desirable to foreclose future bad actions -- not only to avoid the bad effects that would result thereafter, but also to avoid the ex ante incentive effects arising from the anticipation of such policies, which in such cases -- unlike the base case examined in section 2 -- would be undesirable. A significant complication is that it may not be possible to know in advance which policies to rule out -- which is precisely the reason it is not generally desirable to fix government policy, and one of the reasons

----

[101] For an income tax decrease, ex ante anticipation would be desirable.

[102] As to the latter, see Kaplow (1985a) discussing the parallel between risk (an ex ante perspective) and vertical equity (an ex post perspective). In addition, pure revenue-raising is usually intimately connected to distributional issues -- were it not, head taxes would be used because they are nondistorting.

that it might be thought possible that such policies would be adopted in the future. If clearly undesirable policies could be specified, the question would then be why one should trust one's current judgment more than one's future judgment (here, the "one" might best be thought of as society). This is a large part of what is involved in determining the content of a Constitution, including the determination of how easy it should be to amend.

4.9 Relation to Distribution Policy

This entire investigation has largely ignored traditional questions of distribution (vertical equity). Implicitly, the operating assumption has been that government policy can be segregated into three components: underlying policy, distributional policy, and transition policy (i.e., the process by which changes in other policies should be implemented[103]). The most direct linkage between transition policy and distribution policy arises when the patterns of transition gains and losses is not completely random. As discussed in subsection 2.3.2, it might appear that one's preference for transition relief (including windfall taxation) should be influenced by this pattern.

Yet, in a generalized optimal taxation setting, it would appear that, to the extent one attempted systematically to achieve distributional objectives through transition policy, the effect would be the same as if they were achieved through, for example, the income tax. [See Hylland & Zeckhauser (1979), Shavell (1981).] One difference noted in the earlier discussion concerned the random, rather than systematic incidence of reform, which does favor mitigation on risk grounds (and is the motivation for insurance or dispersed ownership, which tends to make transition relief unnecessary and undesirable). Another possible difference concerns perceptions: to the extent the distributional impact of transitions was sufficiently complex and remote for, e.g., individuals to consider in making labor supply decisions (suppose, to take an extreme, they consider only income as measured in their paycheck),

----

[103] Here, the focus has been on implementing policies other than those aimed at distribution.

it might be that pursuit of distributional objectives through transition policy would offer some possibility of benefit. For many areas of government reform, however, it is not clear that the incidence of gains and losses are such that there are significant systematic effects, and if the prevailing transition policy determined whether to give relief based on its incidence in particular cases, it would be less plausible that private actors would be wholly unaware of the general impact. Finally, if institutional constraints prevented resort to appropriate alternative instruments for controlling the distribution of income, there would be a stronger basis for making case-by-case distributional judgments in assessing transition policy. This issue is essentially that raised in cost-benefit analysis concerning whether costs and benefits should receive distributional weights, depending on their incidence. [See Little & Mirrlees (1974), Mishan (1976).]

5. APPLICATIONS

This section considers a number of important instances of reform: changes in tort liability law; government takings of property; changes in government demand, government contracting, and initial program design; and examples of tax reforms. The purpose is both to illustrate the application of the above analysis, with particular emphasis on the applicability of various qualifications, and to illuminate some important areas of concern. In regard to the latter, it should be understood that all the discussion to follow is heavily qualified. No attempt is made to offer a thorough or definitive resolution of the issues discussed, even in the typical spirit with which much academic discussion takes place. The analysis is intentionally brief, and is limited to offering some of the most direct applications and to presenting suggestions concerning how some factors might be resolved.

5.1 Tort Liability

A important issue in tort law involves the evolution over the past several decades of the doctrine applicable to hazardous products and substances. One question that arises concerns the extent to which modern, generally stricter doctrine should be applicable to past activities. This subsection explores the issue of retroactive application, beginning with the assumption that changes in doctrine represent the adoption of appropriate rules in light of new information concerning the nature of potential hazards. To take a particular example, some courts,[104] upon observing the long latency periods

----

[104] The same change accomplished through legislation would be subject to much the same analysis. In this respect, it is noteworthy that the tradition is that changes in tort doctrine by common law courts are retroactively applicable to past actions and harms, whereas statutory changes are typically applicable only prospectively. The source of this tradition, in part, is a traditional legal hostility toward retroactivity (explaining the approach to

often involved before harms appear, have modified their interpretations of statutes of limitations such that they do not begin to run until the harm has been discovered, rather than when the harm was initially caused.

This example is one of those used in subsection 3.2 to illustrate the desirability of retroactive application, in addition to providing no relief at the time the reform is announced. The basic argument is that, since the activity has been harmful all along, it is appropriate that all the harm caused be borne by the investor. Considering the legal context, the idea here is that if investors (perhaps upon receiving legal advice) determine, ex ante, that their new venture will somehow be immune from prevailing legal rules, and that those rules will be followed ex post (in terms of no retroactive application and provision of transition relief), they will have no incentive to take potential harms into account. On the other hand, if it is anticipated that the rule will be not only be changed upon discovery of the harm, but made fully applicable to all the effects of the investor's decisions from the beginning, investment will efficiently take into account the appropriate risks.

As discussed in note 104, such modifications are typically made retroactive when announced by courts, but not when enacted by legislatures. Interestingly, despite controversy over the desirability of such retroactive application, few dispute the appropriateness of applying the new rules as to *future* harms from *past* investments. In other words, it is generally agreed that conventional transition relief is undesirable, although the implicit relief, as described in subsection 3.2, entailed in failing to make retroactive application is often viewed differently. This distinction is often grounded on the assumption that future behavior can be controlled whereas past behavior is sunk. This is wrong on both counts, as the latter argument ignores the ex ante incentive effects and the former ignores that there can be substantial transition losses because, once investment is in

----

legislation) combined with the view dating at least to Blackstone that courts do not "make" new law, but merely "find" what has always been there (producing the notion that since the "new" rule was always the law, application to past acts follows automatically, and does not entail retroactivity). [Traynor (1977).]

place, future limits on its usefulness or taxes on its use can substantially decrease its value.[105]

The ex ante considerations relate to inquiries often made into whether the investor knew of the harm at the time of the initial decision. Making liability turn on such an analysis has two problems, in addition to the direct fact-finding difficulties that arise years or decades after the fact, when records may also have been destroyed. First, such knowledge is, more realistically, probabilistic; hence, the all-or-nothing inquiry into culpability misses much of the reality of the situation. If investors are liable for all the harm that actually results, the proper ex ante incentives will be created. Considerations of risk -- particularly taking into account prospects of bankruptcy -- can complicate the situation. One could view such inquiries as after-the-fact attempts to assess the efficiency of the original decision. If such assessments are accurate, proper incentives will result, so long as the proper ex ante risk analysis is used in making the decision. This is analogous to insurance which makes coverage contingent upon, for example, investment levels or precautions, which can be observed ex post. If such investigations were feasible at reasonable cost, one would expect -- if the legal regime involved retroactive application and there were serious risk-bearing costs (e.g., no dispersed ownership) -- that insurance would be purchased that made coverage contingent upon the determination, e.g., of an arbitration panel, that the investor did not knowingly make inefficiently risky decisions ex ante.

Second, such inquiries often ignore the fact that ex ante knowledge is itself a function of incentives. If future liability arises only if the investor *knew* of the harm with sufficient certainty, the incentive is to forgo investigation and research in making initial decisions. In contrast, if liability is independent of such knowledge, the investor will generally have

----

[105] This is an instance where the future effects highlight the undesirability of direct mitigation or grandfathering in contrast to compensation. As to future behavior (e.g., level of output and shut-down decision), compensation is lump sum -- unless, because of difficulties of valuation, it is based, e.g., upon future output.

an appropriate incentive to discover relevant information concerning possible risks. In principle, an ex post investigation could make liability contingent upon inefficient decisions in this context, although that seems even more difficult than the direct inquiry into prior knowledge.

This case can usefully be contrasted with market risks involved in manufacturing. Just as a product may later be banned (or made prohibitively costly to sell) because it turns out to be dangerous, so a competitor may develop a superior product that makes one's earlier investment obsolete. Firm's also often produce with the risk that they will later be found liable for patent infringement -- which may not only decrease the value of sunk investments, but also may require damage payments for past output. A priori, there is no clear distinction between these cases of market risk and uncertainty in future liability in terms of risk bearing and incentives that justify a general rule of relief for the latter, but not the former.

This analysis assumes, like the base case presentation, that the change in tort doctrine was desirable. If the change was undesirable, it clearly is the case that relief (ideally, an infinite delay in implementation) would be beneficial. Part of the point of this investigation is to emphasize the extent to which this argument for transition relief is best understood as a second best attempt to inhibit the reform when, for some reason, effective direct opposition is infeasible.

One might ask, however, whether a practice of providing transition relief is likely to improve the decisionmaking process of courts or legislatures in this area. Neither body directly bears the costs or benefits of its decisions. In particular, courts do not really have access to budgets from which to pay relief, which would force them to forgo prospective application to provide relief -- an alternative less desirable than compensation, for reasons that are obvious in this example. If courts are reasonably viewed as disinterested decisionmakers, one might find the current practice quite adequate in this regard. When changes are under the control of legislators, the discussion concerning political feasibility, noted earlier, would be applicable. Since, except in the constitutional area, legislatures can override reforms instituted by courts, one is not surprised that many current battles concerning liability for hazardous products and substances are making their way to the legislature, often in the form of proposals that substantially curtail liability for past actions. The fact that courts may in fact be disinterested, largely immune from many forms of interest group activity, may explain part of what is currently observed. On the other hand, it could be contended the differences in expertise between these institutions are the central explanation or justification for the shift to the legislature. This ongoing dispute highlights the extent to which an intelligible discussion of the appropriateness of transition relief must be grounded in some assumption concerning the underlying government policies.[106]

5.2 Takings

The government must compensate for land that it takes for public projects. Absent complicating factors, such compensation will have the adverse impacts described throughout this investigation, which would suggest that current policy is inefficient.[107] But a number of the necessary assumptions may fail

----

[106] In regard to some of the other institutional considerations, it should be noted that the tradition that court decisions are to be broadly applicable to similar cases tends to reduce abuse, minimizing the need for a general rule against retroactive application in this context to accomplish such objectives. In addition, in terms of the credibility of transition policy, courts have proven themselves capable of adhering to practices such that certain traditions seem extremely likely to persist in the future. The analysis presented here is relevant in considering which sorts of practices should be made into traditions and which are best left more open, to respond to new information and changes in circumstances that arise.

[107] To avoid possible confusion in this context, it is worth considering the distinction provided by the ex ante incentives argument between takings (where compensation distorts incentives) and government purchases (where payment is necessary to maintain incentives). Consider the case where the government must take some factories to obtain land for a highway, and also must purchase some cement for the construction. It is not the case that the incentives argument against compensation for takings suggests that the government should not pay for the cement. The reason is simply that, if it was anticipated that cement needed by the government would be taken without payment, there would be a disincentive to produce cement, and the disincentive would be greater the greater the government's needs were expected to be, which is precisely the opposite of the desired incentive effect. When the government pays the market price, which might itself be elevated by the government's demand if that demand is significant relative to the market, the proper ex ante incentives result. The difference with the factories is that they are worthless if the highway is to be built. (Note that it is generally irrelevant whether the government compensates for the land, which is useful, but not generally affected by prospects of future payment. In the extreme case where the land

to hold in this context, making the ultimate conclusion far more uncertain.

Initially, consider the most relevant market imperfections. First, takings will often be low probability events, suggesting that the transaction costs of insurance arrangements (when dispersed ownership does not deal with risk) may be a serious inhibiting factor, since they must be borne ex ante, and thus in all states. On the other hand, one could imagine, e.g., for homeowners, that takings would be included in their general policies, or available as a simple rider. If such provision is thought to be adequate for earthquake protection, it is not clear that takings would be any different.

Second, the probability misperceptions problem could be serious in this context -- again, particularly for homeowners. As discussed previously, the optimal solution here is generally compulsory insurance. If this were a separate program, one could imaging that, in combination with the transaction cost problem, simple compensation would be the efficient solution in this context. On the other hand, as the probability of takings rise -- e.g., as a highway project becomes a realistic possibility (say, 10% for a given parcel of land), the balance would eventually change, making insurance the preferred alternative. (And, as the probability became significant, serious misperceptions would tend to be less problematic.) Thus, the optimal solution might involve compensation in general, with a switch to compulsory insurance (premiums could be added to property tax bills) once particular prospects became serious. This switch is more problematic than it might appear because there would be no insurance for the risk imposed by the new premiums, which would rise as a project became more likely. To deal with this problem, one could have such premiums apply only to investment that is made after the shift.[108] One value of relying on the market in a changing climate such as

----

was created, leveled, or filled in a manner making it more useful for future government projects, then compensation for such improvements should in fact be made.) For further discussion of issues that arise in the takings context, see Kaplow (1986).

[108] For any investment, the premium for the future would be that which was in effect at the time the investment was made. This raises many complications, including when, if ever, the premiums ever end -- e.g., do they continue after a project is canceled? This should be solved by having the premium be a one-time payment for the future risk, as assessed at the time of the investment. There may also be problems in measuring the amount of new investment.

this is that it will automatically begin to offer insurance as the probabilities rise. Such a response, of course, can still be inferior because of misperceptions of risk, but the transaction cost consideration would be less important, unless the change was sudden.

A perhaps stronger argument for compensation in this context can be based on some of the issues discussed in subsection 4.7, which relate to the assumption that the underlying policy is desirable. As noted previously, even if this assumption is violated, compensation will still be inefficient assuming that the land is still to be taken. Thus, the most natural argument would concern the incentives of the government to build the project that requires the taking. The fiscal illusion argument, discussed in subsection 4.7.1, has, in fact, most often been raised in connection with takings, although the previous discussion indicated that the argument hardly provides a clear justification for compensation. It may be, however, that the feasibility of desirable projects is more at issue in the absence of compensation, as it was previously noted that losers tend to be particularly concentrated and thus are more likely to exercise the necessary political influence to block beneficial action unless, as discussed in subsection 4.7.2, they are paid off. Since a taking imposes such substantial losses, it seems doubtful that the logrolling alternative will be feasible in this context. If this factor is most important, one would expect that the government would often want to offer substantial compensation, even if it were not required to do so, as is currently the case.

This discussion, however, overlooks that, in the absence of a requirement of compensation, property owners might be expected to purchase insurance. Moreover, since insurance premiums can be made a function of value (as with fire or other homeowners insurance, greater coverage costs more), moral hazard may not be much of a problem, which could well produce full coverage. (Of course, as discussed earlier in this subsection, transaction costs or other imperfections might impede this result, calling for additional measures.) In that event, the argument concerning pressure groups would change, as the relevant lobbyists would be the insurance companies. At the local level, with insurance companies outside the jurisdiction, one might expect the pressure to be less than that resulting from homeowners who had most or all of their assets at stake -- although one could imagine a number of other factors that could produce a different outcome. In addition, when the property owners consist in large part of widely-owned firms, these political considerations could also be different.

Finally, it should be recalled from subsection 4.7.4 that compensation might be defended as a means to limit abuses of power -- in the extreme case, punitive takings designed to punish a particular individual or unpopular minority. In fact, the historical evidence suggests that such concerns may be the best explanation for the existence of the compensation requirement. [Kaplow (1986).] Of course, if individuals were fully insured -- an option not likely to be taken seriously two hundred years ago -- the ability to use compensation punitively would be largely unaffected by whether or not there existed a requirement of compensation . It should be noted that this consideration, like the fiscal illusion argument but unlike the political feasibility argument, indicates the need for an externally enforced requirement of compensation, rather than making it available at the decisionmaking body's discretion.

Despite the ex ante incentives argument, there seem to be many reasons one might favor a policy of compensation for takings -- far more basis than appeared to exist in the case of changing tort doctrine, where transition relief is not observed, particularly when reforms are announced by courts rather than enacted by legislatures. Yet none of the many arguments makes a decisive case, and it is conceivable that compensation in many contexts does more harm than good, suggesting that the general acceptance of the current approach should be examined more seriously.

5.3 Changes in Government Demand, Government Contracting, and Initial Program Design

A large portion of government activity concerns its demands for goods and services; correspondingly, a large portion of government-created risk arises due to uncertainty concerning its future level of demand. In fact, major changes in government budget allocations have perhaps been as important in this regard as changes in governing rules, including the tax system. Subsection 2.3.1 offered demand uncertainty as an illustration of the similarity of issues created by government and market risk. Based upon the ex ante incentive analysis, one would expect that transition relief would be undesirable in this context as well.[109] In fact, one does not generally observe the government offering compensation (or engaging in windfall taxation) when changing circumstances lead it to reduce (increase) its demand.[110] Other forms of relief would correspond to continuing a previous demand level for some period of time, which may result from political pressures, but is not generally viewed as an appropriate response to reliance of investors on the government's demand.

It is possible, however, that protection against policy change is implicitly incorporated into initial program design -- in this context through long-term contracting. Clearly, if the government is committed for extended periods of time, a later need to change course will be frustrated. In

----

[109] An argument to the contrary, based upon violation of the assumption of optimal government behavior, would be that without any mitigation requirement, the government might behave strategically, beginning new initiatives at a far higher level than that at which they will be continued, to induce excessive private investment, thus lowering prices for future procurement. One might question the ability of our government processes to implement such strategic action, which requires substantial secrecy as to ultimate objectives. One possible exception is noted below.

[110] The discussion in subsection 3.2 suggested that full retroactive application was often desirable when reform was motivated by new information suggesting that a different course would have been desirable from the outset. The exception, illustrated by the takings example, was where the ex ante change in private activity would not have been desirable given that the government did not act earlier. This latter characterization will often be at least partly applicable to changes in government demand, as the government will have already used the goods and services, and the price paid will reflect a market equilibrium that reflects the marginal value of such goods and services for nongovernment uses.

essence, such contractual commitments translate into delayed implementation of future reforms that might become desirable -- the amount of delay being determined by the length of the contract. Of course, an alternative is to enter shorter contracts, or specify cancellation provisions -- e.g., indicating that the contract may be nullified if the program is terminated, but not otherwise.

Much government demand is met using very short-term contracts, most commonly spot purchasing, at-will employment, or one-year arrangements with private entities.[111] Exceptions might exist for a complex set of reasons. Civil service employment -- which protects not only individual employees, but also limits the ability to quickly shift budget priorities -- is often defended on the ground of preventing government abuses. A more significant example for present purposes involves long-term, specialized procurement arrangements, such as are common in the defense sector. (Even these contracts are often in stages, with orders for large numbers of a weapons system not being made final until after successful testing or delivery of a few items, and quantities often being set at significantly different levels than originally contemplated.)

Such long-term contracts might be defended on the ground that the government is better able to absorb the large risks that are involved, although prior discussion has cast doubt on the possibility that the government could significantly outperform private arrangements. Of course, to the extent moral hazard problems -- often noted in defense contracting in connection with cost-plus arrangements -- are the reason private risk mitigation is incomplete, the analysis suggests that government mitigation will only make matters worse.

An alternative possibility is that the government has better information. Subsection 4.3 suggested that this generally would not be the case, and that

----

[111] Formally, most government demand, unlike legal rules, has an implicit one-year sunset provision, since annual appropriations are required. Yet many projects are long-term and budgetary reform is often incremental, so future funding in general may have a range of predictability to that for legal rules.

when it was, the government could simply communicate the information to the relevant private actors. One could imagine, however, that with complex defense systems, some of the information would be costly to communicate (a problem worsened by security concerns) and the possibility of strategic misrepresentation (see note 109) might be more severe. For purposes of determining ex ante investment levels, it might be more accurate to view the government rather than the private sector as the relevant decisionmaker: decentralization is forgone in this context.[112] So long as the government has made the proper ex ante calculus, taking into account the probabilities that future information or circumstances will affect the ultimate value of the project, there is no conflict between this approach and the argument presented in this investigation.

A related example concerns initial program design outside the contracting context. Often subsidies are provided as a single payment (or with significant front-loading) at the time an investment is made (or shortly thereafter) rather than when the flow of benefits is produced. For example, most tax credits -- such as for investment generally (ITC) and special provisions for energy conservation -- provide an approximately instantaneous subsidy. In the event of a later repeal -- e.g., of conservation subsidies, because the energy crisis proved short-lived -- there is an implicit grandfathering, because if the subsidy had been linked to the actual benefits, repeal without relief would have resulted in no future subsidy, whereas the subsidy corresponding to future periods has already been paid up front in the cases under consideration. To avoid grandfathering, the appropriate ex post action would be a one-shot tax to take back the amount of the initial subsidy properly attributable to the future (defined from the date of the reform).

----

[112] Significant decentralization might consist of private firms creating and building weapons systems, marketing them to the government, if it is interested, upon completion. Problems related to small numbers bargaining, information concerning demand (especially in light of security interests), and coordination, among others, suggest why this option is not used, although, for purposes of government demand by smaller countries, this system does (loosely) describe reality, with producing countries often playing the role of competitive firms.

This is another instance where retroactive action, as discussed in subsection 3.2, is necessary to avoid mitigation.

In principle, such implicit grandfathering can be justified in two ways. First, if the level of the one-shot subsidy takes into account the future contingencies, rather than being based an the correction that would appear appropriate if the current situation continued indefinitely into the future, an appropriate ex ante incentive results, and risk is avoided. (The analogy is to full insurance when investment is observable.) Second, if the benefits that are the object of the subsidy result from the act of investment per se, rather than the future flow of activity produced by the investment, a one-shot provision is appropriate, and no implicit grandfathering would be involved. This characterization seems inapplicable in the energy conservation situation; for the ITC, the issue depends upon the justification for the credit. If neither explanation is applicable, an ex post tax is necessary to avoid the undesirable effects of grandfathering.

The first justification -- where the level of the ex ante provision already accounts for future contingencies, as was also noted as a possibility in the defense contracting context -- is of general application. Thus, for example, no investment distortion would result from full compensation for takings if there were an appropriate ex ante tax (i.e., compulsory insurance, with the tax being the premium), as discussed in subsection 5.2. Again, the analogy is to the case where investment is observable by the insurance company, and a first best can therefore be achieved. In other words, such a government approach, broadly considered, constitutes a system of general compulsory insurance for government risks. To preserve the appropriate incentive features, it is important that the premium be charged, and made a function of investment levels where possible, and that in other cases supplemental private insurance be banned (to prevent the two-stage moral hazard problem).

The general feasibility of such an approach will vary greatly by context, depending particularly on available information, as well as administrative and transaction cost characteristics. For a wide variety of complex risks, in areas where dispersed ownership of affected investments exists, one suspects that such an approach would be undesirable. And in cases where the relevant investment levels and actions cannot be observed, such an approach is impossible. (If investment is observable, private arrangements can yield a first best without government relief, so long as the information available to insurance companies is as good as the government's. On the other hand, one can imagine examples such as energy conservation by homeowners, an ex ante one-shot subsidy reflecting the present value of social benefits, taking into account the prospects of future change, seems administratively the simplest, and has the added benefit that, to the extent consumers are myopic or constrained in borrowing, better investment decisions might result than if private risk-mitigation were used instead.

Finally, it is interesting to apply this discussion to the frequently-voiced argument that transition relief -- e.g., grandfathering in the event that the municipal bond interest exemption is repealed -- is appropriate, because private investors have essentially "contracted with the government," the implicit terms being continuation of existing government policy. But no one rationally expects, or is an any way entitled to expect that the government will not change its policies when new information or changes in circumstances make reform desirable. Moreover, it should be clear that this formulation is circular. [See Kaplow (1986).] If it is assumed that relief in the event of reform was part of the deal, providing relief would seem appropriate. But one could equally argue that the contrary assumption is correct (after all, the contract's relevant terms were implicit), in which case relief would affirmatively be inappropriate. This investigation has focused on which assumption, if understood and adhered to in the long run, results in the most efficient outcome. The general argument has been against relief. This subsection suggests an important caveat when government policy can be and is initially tailored with the future contingencies in mind,[113] and

----

[113] In many instances, it might be most natural to say that the government has a "null" policy -- e.g., many industries are not regulated, many activities do not receive unique differential taxation, etc. Thus, when significant changes do occur in such contexts, it often will be implausible to

suggests factors relevant to determining when such action is appropriate.

5.4 Tax Reform

In the past decade, major tax reform appears to have become a way of life. Even before then, substantial reform occurred on a regular basis. Since the possible bases for tax policy, and thus tax reform, are so varied, it is necessary to consider particular contexts and issues separately. For example, substantially different questions are raised by an increase in rates to raise further revenues, a shift from an income tax to a consumption tax, the closing of loopholes, changing the progressivity of the system, and altering a Pigouvian tax. The latter case is closest to the base case used throughout this investigation, and most clearly illustrates the thesis advanced here. This subsection briefly considers two issues that have arisen in connection with major proposals for tax reform.

5.4.1 Transition from an Income Tax to a Consumption Tax: Carryover Problems versus Price Changes

Discussion of major tax reform, particularly that addressing the transition from an income tax to a consumption tax, has devoted substantial attention to the appropriate transition treatment of carryover problems and price changes -- and, in particular, to whether different treatment is warranted.[114] Carryover problems involve effects caused by that might best be characterized as changes in accounting rules.[115] For example, the income tax applies to income that is saved; when an asset is sold, the taxpayer deducts

----

believe that the previously existing government policy was designed with such contingencies in mind.

[114] Some details of the debate are discussed in Kaplow (1986).

[115] Calling them accounting rules is not to argue that changes may not be of great substantive importance. In particular, accounting practices determine timing, which is of critical importance in light of the time value of money, and is at the core of arguments over whether there would be an income tax or a consumption tax.

the basis when computing gain, in order to avoid double taxation. Under a consumption tax, there is no initial taxation of income that is saved; therefore, when an asset is sold, the entire proceeds are included in the tax base. As a result, for an asset purchased under an income tax regime and sold after transition to a consumption tax, double taxation would result if the transition rules did not permit taxpayers to take into account their income tax basis. In contrast, price change effects are those akin to that caused by repeal of the oil depletion allowance or other special deductions and exclusions. Most proposals for a shift from an income tax to a consumption tax would alter the relative treatment of many transactions. Of course, it is generally possible to formulate a consumption tax proposal that differs from the current income tax (or any other) only in its core provisions, preserving special treatments where they exist.

It should not be surprising that the appropriate transition treatment of carryovers and price changes has proved to be difficult and controversial since commentary on the subject does not offer any systematic framework for examining transition questions. For example, some commentators note that both should be treated differently because each has different causes, and others advocate the same treatment because, after all, investors care only about how much they gain or lose, and not ultimately about the cause. The approach presented here argues for different treatment: in the simple case, transition relief should be available in the case of carryovers and not in the case of price changes. Begin with the former.

Consider the simple example of a mandated accounting change from calendar-year accounting to fiscal-year accounting -- for example, July 1 to June 30. In the absence of transition rules, six months of net income would either be taxed twice or exempted, depending upon how the change was made. The obvious solution is to construct a "short year" to deal with the six-month transition period, just as the current income tax system generally does for individual taxpayers who change their accounting period. Transition relief (to avoid double taxation -- or transition penalties, to avoid exemption) is appropriate because it is necessary to measure the desired tax base. In making the transition from an income tax to a consumption tax, failing to make adjustments to basis (and related adjustments for other carryovers) would not serve the purposes of the consumption tax, because none of the principles that define its tax base call for the effects on lifetime tax burdens that would result from avoiding transition relief.[116]

Contrast such accounting issues with price changes. Price changes are in fact the standard sort of example used throughout this investigation, where ex ante incentive effects suggest that transition relief is undesirable. Thus, for example, if a deduction or exemption is repealed because it is no longer the case that the social benefits of an activity exceed its private benefits, the ex ante effect of no relief, which would involve curtailing investment, would be desirable. In contrast, in the simple example of the mandated change in accounting year, the ex ante effect would be to virtually cease economic activity in the six-month period, if it were to be double-taxed, or to shift large amounts of activity into that period, if it were to be exempt. As in the case with the transition to a consumption tax, such changes in behavior are not desirable from perspective suggested by the motivation for the reform. Therefore, there does seem to be a solid basis for distinguishing these two effects in fashioning tax transitions.

5.4.2 Lowering Tax Rates, the ITC, and ACRS Windfall Recapture

The central feature of current tax reform is a significant lowering of rates for both the individual and corporate income taxes. One of the major accompanying changes that helps preserve revenue neutrality is the repeal of

----

[116] In fact, failing to provide such relief would be counterproductive from the consumption tax perspective, as it is frequently discussed, because the ex ante effects of anticipating the reform, if it is not to be accompanied by carryover adjustments for basis, would be to decrease savings incentives. It seems possible that, in present value terms, the ex ante decrease could be more significant that the increase arising after adoption -- although there may be significant catch-up savings after the effective date, and pre-enactment consumption might involve real savings through the purchase of consumer durables, depending on how such assets are treated in the transition.

the ITC. This combination is found objectionable by many economists because is runs directly contrary to the confiscatory tax prescription, discussed briefly in subsection 4.8, because the result is substantially higher effective marginal rates for new investment (causing a negative incentive effect) and substantially lower rates for old investment (which is already in place, and thus produces windfalls with no positive incentive effects).[117] Here, the credibility problem is reversed: with a confiscatory approach, the problem is the fear that it will be repeated; with this reverse approach, it is difficult to understand from this perspective why it would be done in the first place.

The proposal for ACRS windfall recapture provides an interesting contrast. The argument in its favor has been that ACRS provided rather generous treatment of new investment to begin with. Part of the implicit bargain (compare the discussion in subsection 5.3) was that income produced by such projects in later years would be taxed at the then-prevailing 46% marginal rate.[118] With a decrease in the marginal corporate rate to 33%, a windfall is conveyed to such investors, so it is suggested that a windfall tax be imposed to avoid this transition gain. The logic for investment incentives is clear. Since the investment is already in place, this added tax cannot hurt incentives, and it allows the government to finance lower marginal rates for the future. The arguments in favor of this provision are essentially those opposing lower rates financed by repeal of the ITC.

What should be clear is that the logic supporting the ACRS windfall recapture approach extends well beyond that context. Most obviously, it applies to *any* prior investment that is not to be subjected to a lower rate -- including pre-ACRS investments, and investments in human capital that will, under the current reforms, be taxed at lower marginal rates. Moreover, if

----

117 Another proposal not currently at the center of public attention raising similar issues involves the integration of the corporate income tax into the individual income tax. Many such proposals attempt to limit the benefits to new equity precisely for the reasons discussed in text.

118 For individuals with such property, the relevant marginal rate would often be 50%, and in any event, individual rates are also to be lowered across the board.

such past investment is viewed as sunk, with no harm to result from a windfall recapture, why not tax it at a full 100% of current value, rather than merely making up the difference between the old and new rates? The logic of the argument suggests that there are no costs to such windfall taxation, and benefits rise with the amount of revenue raised.

As noted in subsection 4.8, it has long been known that these latter statements are correct from a purely ex post perspective, if future credibility is not a problem. This investigation suggests that an ex ante perspective further illuminates such issues. Subsection 4.8 emphasized how transitions relating to raising revenue are different from other transitions because of the opposite preference concerning the desirability of ex ante anticipation. In the case of tax *decreases*, however, ex ante anticipation would have desirable effects, if the tax distorts incentives. Thus, the prospect of lower rates in the late 1980's might have encouraged investment earlier in the decade. Those earlier years have passed, but for purposes of future changes, it is not obvious that the best long-run transition policy involves no transition relief for losses caused by tax increases (as is the general practice[119]) and full windfall taxation for gains resulting from tax decreases (or, equivalently, not enacting general decreases that apply to old investment in the first place[120]). This is an ad hoc compromise, and its desirability is likely to be sensitive to the appropriate discount rate and views concerning the credibility of announcements as to the future.

This investigation has focused on the appropriate long-run transition policy, under the assumption that transition rules are correctly anticipated and carried out. Subsection 4.6 discussed the difficulties in applying such analysis to a pending reform, but it seems worth considering what sort of transition policy would have been appropriate from an ex ante perspective.

----

119 Interestingly, historical practice has often distinguished changes in rates from changes in particular rules, the former usually being immediately applicable (regardless of the direction of the change) and the latter often involving relief (particularly for unfavorable changes).

120 After all, enactment of an ITC could be viewed equivalently as a rate decrease accompanied by grandfathering of old investment.

Clearly, if a 100% tax on capital had been anticipated, there would have been little long-term investment. Similarly, if lower rates would apply to investment in place, there would be more investment ex ante. On the other hand, the less windfall taxation at the time rates are lowered, the less the rates can be decreased. Prior investigations of the appropriate pattern of tax rates over time -- which in the simplest case would favor level rates because distortions are proportional to the square of tax rates -- are relevant to resolving this issue.

One must consider, however, why it is that a government would need to raise or be able to lower rates at some future time.[121] Suppose, for example, that changes in events relating to foreign policy suggested that a significant long run increase or decrease in the defense budget was appropriate. Once such a change is clearly necessary, tax adjustments can begin (and spread over the future through adjustments in government debt), even if expenditure adjustments are not required for years to come.

But what of the ex ante prospect of such a change when there is significant uncertainty -- e.g., there is a 50% chance that a major adjustment will turn out to be necessary in 10 years? It may be that some immediate, perhaps partial tax adjustment is still appropriate, because of the benefits of more level rates over time, which would raise another round of ex ante considerations relating to the time period when there was uncertainty as to whether this uncertainty would arise. To begin to unravel this regress, consider whether the change 10 years in the future[122] should be accompanied by transition relief.

For concreteness, assume that the change is a major increase in rates. If old investment is protected -- e.g., through grandfathering -- even higher rates on new investment will be necessary, with the result that effective

----

[121] If a change in circumstances only called for a temporary increase or decrease in government expenditures, the need to adjust taxes could be spread over the future by adjusting debt policy, but some change in rates would still be necessary.

[122] Even if some change is made in anticipation, when the event is realized, there will be a later adjustment.

rates over the life of long-run investment projects will greatly differ over time. If the higher rates apply to old investment, there will be less of it before the reform, but less discouragement of new investment afterwards. If the more level pattern is to be preferred, transition relief would seem undesirable. Now consider the prospect of a major decrease in rates, which is closer to the current reforms. If it is known in advance that the lower rates will not be applicable to old investment (see note 120 suggesting that enacting an ITC is much like a grandfathered rate decrease), there will be less investment ex ante, but more than otherwise after the reform, because rates can be lowered further. On the other hand, if all investment will benefit from lower rates, there will be relatively more investment ex ante, but less later because rates cannot be lowered as much. Once again, providing no transition relief (in this case, no windfall tax on gains or grandfathering) provides for more level tax rates on investments over time.

This analysis from the ex ante perspective is only appropriate in the long run, and depends on assumptions and institutional possibilities concerning the credibility of transition policy. Moreover, the discussion was informal, and did not carefully model the effects of taxes on incentives, clearly state the undistorted situation, take into account discounting,[123] or account for second best considerations that can arise when discussing taxation. This is intentional, as consideration of transition issues in the revenue-raising and distribution contexts where not the focus of this investigation. The problem is further complicated by the fact that different taxes and different reasons for changing taxes (e.g., different revenue needs, changing values concerning distribution, changing views concerning the desirability of tax expenditures) will generally affect the analysis. The only clear conclusion is that most current discussion of transition issues in connection with tax reform fails to take an appropriate dynamic perspective.[124]

----

123 Discounting does not obviously favor whichever approach results in greater current investment in preference to future investment, because the effect of future taxes on present investment is less than the effect of similar taxes on future investment, also because of discounting. This is consistent with the notion that level rates over time are best, rather than lower rates and large deficits now, made up by higher rates in the future.

----

[124] Recent attention devoted to changes in future tax policy has not generally considered transition issues of the sort emphasized here. For example, Auerbach and Hines (1986) discuss how tax reform anticipated with certainty affects short-run investment behavior. Poterba's (1984) examination of tax-exempt bond yields relies on the effects of anticipated tax changes to differentiate among competing models. Judd (1984) examines the welfare effect of changes in tax policy depending upon the extent to which they are anticipated. In contrast, this investigation has assumed for the uncertainty case the sort of effects they describe and attempted to consider the efficiency properties -- accounting for both incentives and risk bearing -- of modifying such affects through transition policy. In addition, as noted previously (see page 88), significant attention has also been given to the affect of uncertainty per se on incentives. And Skinner (1986) considers the welfare effects of uncertain tax policy when there is no economic justification for the changes, but does not address market risk mitigation techniques or the efficiency properties of transition policy.

6. CONCLUSION

Government transition policy serves to mitigate the risk imposed by uncertainty concerning future government action. The arguments for and against various forms of mitigation were seen to be quite similar to those arising in the context of market risks, where commonly-held assumptions concerning the limitations of government decisionmaking by comparison to the benefits of a decentralized market process lead to the conclusion that it is best to leave such problems to the market.

The basic reason that government action might be thought necessary is to mitigate risk. Considering this factor alone would support full compensation, although the market often could largely accomplish this result. Adding concerns for incentives complicates the problem because, assuming that compensation (government or private) is anticipated ex ante, it will affect behavior. Generally, less than full relief would be in order and, taking into account market responses, a government policy of no compensation -- or other form of mitigation -- would be appropriate in the simple case.

Alternatively, if the government were to provide transition relief, it is generally preferable for it to offer (possibly compulsory) insurance, charging an appropriate premium. Moreover, to avoid the two-stage moral hazard problem examined here, it would also be necessary to ban supplemental private coverage (or make the government's premium a function of the level of additional coverage). This result has implications for the appropriate form of government insurance generally, such as through provisions for medical care and disaster relief.

Finally, brief note was made of the strong similarity between the case for compensation and that for windfall profits taxation. The only significant distinction is that the former requires that additional revenues be raised whereas the latter diminishes the revenue requirement from costly revenue sources. There may also be some differences in the ability to administer various tax and compensation schemes that would justify different treatment in some contexts.

It is worth considering which of the assumptions necessary to these results might most plausibly be questioned in practice. First, if government policy change over time is not thought to reflect actual changes in which policies are more desirable, the incentives issue would be viewed differently.[125] Second, if the government is not generally constrained to following a consistent transition policy over time -- e.g., if it can convince actors that it will maintain a different policy in the future despite its actions in the past -- the issue of ex ante incentive effects would again be changed substantially. Finally, exceptions to the main argument might be compelling where additional imperfections combine with factors considered here in a manner justifying compensation, as might be the case if incentive problems were believed to be minimal and there is reason to believe that the market cannot adequately spread the risk involved. For example, when the probability of reform is very low, market provision might fail due to transaction costs or the systematic underestimation of future probabilities of reform by some actors. If this third sort of consideration provides the justification for transition relief -- or, in many instances, for compulsory government insurance -- in some settings, it will generally be the case that government risk mitigation of similar market risks will be justified for the same reasons. It may well be that views concerning transition policy contrary to those emerging from the base case analysis are grounded in one or more of these alternative assumptions, which reinforces the need for further investigation of these issues.

----

[125] Ekern (1971) examines the effect of political risk on investors. Making direct normative conclusions based on this analysis would entail the view that political change involves the purely random imposition of risk without serving any social purpose.

Appendix A

Proof of Proposition IV -- Section 2.2.2

Proposition IV: *If $\Delta B > 0$, $d \in (0,1]$, and* U *is a nonincreasing absolute risk aversion utility function, then: $\partial U/\partial\beta|_{0,b} < 0$, where $\beta \equiv d+b|_d$.*

Proof: First, it will be demonstrated that: $\tilde{b}\Delta B'\tilde{K}_b - b\Delta B'K_b > 0$, where $b \equiv b|_d$ and $\tilde{b} \equiv d+b$. Begin by noting from Proposition I that $d \in (0,1]$ implies $\beta \in [0,1]$ and b, $\tilde{b} > 0$, so Lemma III.1 holds for the cases to be examined. Since $d > 0$,

(A1) $\tilde{b} > b$

Recalling that $\beta \in [0,1] \Rightarrow U_{kk}$ strictly negative,

(A2) $\tilde{b}U_{kk} < bU_{kk}$

This step, and a number to follow, assume that expressions are evaluated at the same β, K, and U, which holds by construction and the application of Lemma II.1. Let θ and Ω represent the same expressions as in (18), and perform the following manipulations:

(A3) $\tilde{b}U_{kk} - b\tilde{b}\theta < bU_{kk} - b\tilde{b}\theta$

(A4) $\tilde{b}[U_{kk} - b\theta] < b[U_{kk} - \tilde{b}\theta]$

Using (18), we can make the following substitution:

(A5) $\bar{b}\Omega/K_b < b\Omega/\bar{K}_b$

The possibility that K_b or $\bar{K}_b$ equals zero is ruled out by the unambiguous sign on the right side of (11), as discussed in Lemma III.1. Moreover, since the sign of Ω is the opposite of that of $\Delta B'$, we have:

(A6) $\bar{b}\Delta B'/K_b > b\Delta B'/\bar{K}_b$

Lemma III.1 implies that K_b and $\bar{K}_b$ have the same sign, since $\Delta B'$ is evaluated at the same value of K in both instances. Therefore, there is no change in the inequality by performing the following manipulation:

(A7) $\bar{b}\Delta B'\bar{K}_b > b\Delta B'K_b$

(A8) $\bar{b}\Delta B'\bar{K}_b - b\Delta B'K_b > 0.$

It is now straightforward to demonstrate how this result implies $\partial U/\partial\beta|_{d,b} - \partial U/\partial\beta|_{0,\beta} > 0$. Lemma II.1 indicates that both derivatives are evaluated at the same K and U, in addition to the same β. Simply examining these derivatives (14) and subtracting yields:

(A9) $-b\Delta B'K_b\bar{U}' - (-\bar{b}\Delta B'\bar{K}_b\bar{U}')$, or equivalently

(A10) $\bar{U}'[\bar{b}\Delta B'\bar{K}_b - b\Delta B'K_b]$

$\bar{U}' > 0$, and the sign of the bracketed term is positive.

To complete the proof, simply note from the first order condition for β at an optimum that $\partial U/\partial\beta|_{d,b} = 0$. Therefore, $\partial U/\partial\beta|_{0,\beta} < 0$. *Q.E.D.*

Remark on the Case of d < 0: d < 0 consists of "negative compensation" -- i.e., a tax accentuating the loss. So long as $\beta|_d \geq 0$, the argument of Proposition IV holds, mutatis mutandis (the inequality reverses in (A1), and follows through to the end), with the result that $\partial U/\partial\beta|_{0,\beta} > 0$. For d sufficiently negative, however, it is possible that $\beta|_d < 0$. This disrupts two arguments in the proof. First, the argument in Lemma III.1 that (23) was negative relied on the assumption that $b \geq 0$ in evaluating the second term. Thus, in applying Lemma III.1 to $\bar{b}$ in moving from (A6) to (A7) is problematic. Direct inspection of (11) reveals that $b < 0$ and $d < 0$ imply that all left side bracketed terms are strictly negative except the third, which, if B_o'' has sufficiently greater (negative) magnitude than B_r'', could be positive. If that term is sufficiently positive to dominate the other three, Lemma III.1's conclusion, and thus Proposition IV as applied to $d < 0$, would fail. Similarly, this third term is also the third term in U_{kk}. Therefore, $U_{kk} = 0$ is possible, which would imply $\partial U/\partial \beta|_{0,\beta} = 0$, instead of being positive.

Appendix B

Uncertain Transition Policy -- Section 3.3

B.1 Certain Compensation at the Expected Level of Uncertain Compensation

Designate the states o, r & s, where *o* continues to designate the no reform state and r & s correspond to the reform states. *q* designates the probability of state *s* conditional on reform. α is the proportion of compensation in state *s* -- i.e., α=1 corresponds to compensation of *d* in state *s* and no compensation in state *r*, α=0 corresponds to compensation of *d* in state *r* and none in *s*. The utility function to be maximized is:

$$(*A1)\quad U = (1-p)U[B_o - c] + p(1-q)U[B_r - c + \frac{1-\alpha}{1-q}d\Delta B] + pqU[B_r - c + \frac{\alpha}{q}d\Delta B]$$

where

$$(*A2)\quad c = pd[(1-q)\frac{1-\alpha}{1-q} + q\frac{\alpha}{q}]\Delta B = pd\Delta B$$

The first order condition for *K* is

$$(*A3)\quad \frac{\partial U}{\partial K} = (1-p)U'_o + p(1-q)U'_r[B'_r + \frac{1-\alpha}{1-q}d\Delta B'] + pqU'_s[B'_r + \frac{\alpha}{q}d\Delta B'] = 0$$

Note that B'_r is also used for state *s* since the only difference between states r & s is the amount of compensation offered. For α, the first order condition is

$$(*A4)\ \frac{\partial U}{\partial \alpha} = (1-p)U_o'[B_o'K_\alpha - pd\Delta B'K_\alpha]$$

$$+ p(1-q)U_r'[B_r'K_\alpha - pd\Delta B'K_\alpha + \frac{1-\alpha}{1-q}d\Delta B'K_\alpha - \frac{1}{1-q}d\Delta B]$$

$$+ pqU_s'[B_r'K_\alpha - pd\Delta B'K_\alpha + \frac{\alpha}{q}d\Delta B'K_\alpha + \frac{1}{q}d\Delta B] = 0$$

Using (*A3) to eliminate terms from (*A4) and regrouping, one has

$$(*A5)\ pd[\Delta B(U_s' - U_r') - \Delta B'K_\alpha \bar{U}'] = 0$$

The first term is the risk spreading effect, and is positive when there is less income in state *s*, which corresponds to $\alpha < q$ (and conversely for $\alpha > q$). The second term is the moral hazard effect. For (*A5) to hold, it must be that (assuming the basic case, where $\Delta B > 0$ and $\Delta B' > 0$) $K_\alpha > 0$ in this range (and the sign reverses when $\alpha > q$). Similarly, for $\alpha=q$ (which corresponds to equalization across states r & s) to be an optimum, it must be that $K_\alpha=0$.

This latter result can readily be demonstrated to hold. Taking the derivative of (*A3) with respect to α and grouping terms yields:

$$(*A6)\ K_\alpha\left[U_{kk} - pd\Delta B'[(1-p)B_o'U_o'' + p(1-q)(B_r' + \frac{1-\alpha}{1-q}d\Delta B')U_r'' + pq(B_r' + \frac{\alpha}{q}d\Delta B')U_s''\right]$$

$$= -pd\Delta B'(U_s' - U_r') - pd\Delta B[(B_r' + \frac{\alpha}{q}d\Delta B')U_s'' - (B_r' + \frac{1-\alpha}{1-q}d\Delta B')U_r'']$$

At $\alpha=q$, one also has $1-\alpha=1-q$ and $Y_r=Y_s$. Both right side terms are thus equal to zero. On the left side, if $\Delta B' > 0$, the inner bracketed term is nonnegative so long as absolute risk aversion is nonincreasing. (This can be demonstrated by comparing this term with (*A3), and noting that, with $\Delta B' > 0$, the bracketed term here puts more weight on the negative terms, but since all the weights are negative, the net must be positive. For $\Delta B' < 0$, the sign of the bracketed term reverses, preserving the conclusion to follow.) $d \in [0,1]$ is sufficient to guarantee that U_{kk} is strictly negative, as in the base case. Therefore, $K_\alpha=0$. Finally, upon examining (*A6), where is nothing apparent that rules out the required sign for K_α to permit a local extreme point other than at $\alpha=q$.

B.2 Random Resolution of Uncertainty Ex Ante

For this alternative, it is even more difficult to reach determinate results. As a result, the discussion here will be confined to stating formally the maximization problem to clarify any ambiguity in the textual discussion.

For this case, states r & s are again two "reform" states, *r* with no compensation and *s* with compensation of *d*. *q*, in addition to being the probability of compensation in the original problem with uncertain transition relief, also represents the probability that the ex ante resolution is favorable to compensation or, equivalently, the proportion of identical projects that will receive compensation (with certainty). The new variables, ρ & σ, represent the probability of state *s*, *contingent on adverse resolution or on being in the subset that receives no compensation* and *contingent on favorable resolution or being in the subset that receives compensation of* d, respectively. In addition, the weighted sum of ρ & σ is *q*.

The utility function is thus

$$(*A7)\ U = (1-q)[(1-p)U(B_o - c) + p(1-\rho)U(B_r - c) + p\rho U(B_r - c + d\Delta B)]$$
$$q[(1-p)U(B_o - c) + p(1-\sigma)U(B_r - c) + p\sigma U(B_r - c + d\Delta B)]$$

It is also required that

$$(*A8)\ q\sigma + (1-q)\rho = q, \text{ or}$$

$$(*A9)\ \sigma = 1 - \frac{1-q}{q}\rho$$

For break-even, the constraint is

$$(*A10)\ c = pd[\rho\Delta B(K^1) + \sigma\Delta B(K^2)]$$

where K^1 is the investment level that maximizes the first line of (*A7) and K^2 is that which maximizes the second line of (*A7).

The interpretation is as follows. For 1-q percent of the projects (or for all projects, with probability 1-q), the first line of (*A7) is maximized;

otherwise, the second line is maximized. Consider the case where $\rho=\sigma=q$. In this case, the problem reduces the pure case of maximization in light of uncertain transition relief. The case where $\rho=0$ and $\sigma=1$ corresponds to the pure forms of the alternatives discussed in text. For 1-q of the projects (or for all, with probability 1-q), there is no relief, and for q of the projects (or for all, with probability 1-q), there is compensation of *d* with certainty. The uncertainty depicted by *q*, under the probabilistic interpretation, is resolved *before* investment is made, so either interpretation of the prospect produces the same K^1 and K^2, and thus the same expected utility.

The first order conditions for maximization over K^1 and K^2 are straightforward. As ρ is reduced from *q* to zero, there is less moral hazard with respect to K^1 and more with respect to K^2. Examining the first order condition for ρ proved generally unenlightening, except that it is clear that, at $\rho=\sigma=q$, $\partial U/\partial \rho$ can have any sign. (It is zero at q=.5, although one suspects this to be a local minimum rather than a maximum.)

REFERENCES

Akerlof, G., The Market for Lemons: Quality Uncertainty and the Market Mechanism, Quarterly Journal of Economics, 84, 488-500.

Arrow, K.J., 1963, Uncertainty and the Welfare Economics of Medical Care, American Economic Review 53, 941-74.

Arrow, K.J. and R.C. Lind, 1970, Uncertainty and the Evaluation of Public Investment Decisions, American Economic Review, 364-78.

Auerbach, A.J. and J.R. Hines, 1986, Tax Reform, Investment, and the Value of the Firm, NBER Working Paper 1803.

Blume, L. and D.L. Rubinfeld, 1984, Compensation for Takings: An Economic Analysis, California Law Review 72, 569-628.

Blume. L., D.L. Rubinfeld, and P. Shapiro, 1984, The Taking of Land: When Should Compensation be Paid?, Quarterly Journal of Economics 99, 71-92.

Brennan, G., 1971, Horizontal Equity: An Extension of an Extension, Public Finance 26, 437-56.

Dahlby, B.G., 1981, Adverse Selection and Pareto Improvements through Compulsory Insurance, Public Choice, 37, 547-xxx.

Downs, A., 1957, An Economic Theory of Democracy.

Ekern, S., 1971, Taxation, Political Risk and Portfolio Selection, Economica 38, 421-430.

Feldstein, M., 1976a, On the Theory of Tax Reform, Journal of Public Economics, 6, 77-104.

Feldstein, M., 1976b, Compensation in Tax Reform, National Tax Journal 29, 123-30.

Foldes, L.P. and R. Rees, 1977, A Note on the Arrow-Lind Theorem, American Economic Review 67, 188-193.

Grossman, H., 1979, Adverse Selection, Dissembling, and Competitive Equilibrium, Bell Journal of Economics, 10, 336-343.

Grossman, S. and O. Hart, 1983, An analysis of the Principal-Agent Problem, Econometrica, 51, 7-45.

Hartman, R., 1985, Uncertainty in Future Government Spending and Investment, Quarterly Journal of Economics, 100, 1339-1347.

Holmstrom, B., 1979, Moral Hazard and Observability, Bell Journal of Economics, 10, 74-91.

Hylland, A. and R. Zeckhauser, 1979, Distributional Objectives Should Affect Taxes But Not Program Choice or Design, Scandinavian Journal of Economics, 264-84.

Jensen, M.C. and W.H. Meckling, 1976, Theory of the Firm: Managerial Behavior, Agency Costs, and Ownership Structure, Journal of Financial Economics, 305-360.

Judd, K.L., 1984, The Welfare Costs of Factor Taxation in a Perfect Foresight Model, mimeo.

Kahneman, D., P. Slovic, and A. Tversky, eds., 1982, Judgment Under Uncertainty: Heuristics and Biases.

Kaplow, L., 1985a, Horizontal Equity: Measures in Search of a Principle, NBER Working Paper 1679.

Kaplow, L., 1985b, Optimal Government Policy Toward Risk Imposed By Uncertainty Concerning Future Government Action (mimeo, earlier version).

Kaplow, L., 1986, An Economic Analysis of Legal Transitions, Harvard Law Review 99, 509-617.

King, M.A., 1983, An Index of Inequality: With Applications to Horizontal Equity and Social Mobility, Econometrica 51, 99-115.

Kitch, E.W., 1977, Can We Buy Our Way Out of Harmful Regulation?, in: D.L. Martin and W.F. Schwartz, Deregulating American Industry: Legal and Economic Problems (Lexington Books, Lexington, Mass.).
Kunreuther, H., et al., 1978, Disaster Insurance Protection: Public Policy Lessons (Wiley, New York).
Kunreuther, H., W. Sanderson and R. Vetschera, 1985, A Behavioral Model of the Adoption of Protective Activities, Journal of Economic Behavior and Organization, 6, 1-15.
Kydland, F.E. and E.C. Prescott, 1977, Rules Rather than Discretion: The Inconsistency of Optimal Plans, Journal of Political Economy 85, 473-491.
Lind, R.C., 1982, A Primer on the Major Issues Relating to the Discount Rate for Evaluating National Energy Options, in: R.C. Lind, et al, Discounting for Time and Risk in Energy Policy (Resources for the Future, Washington), 21-94.
Little, I.M.D. and J.A. Mirrlees, 1974, Project Appraisal and Planning for Developing Countries (Basic Books, New York).
Mayers, D. and C.W. Smith, 1982, On the Corporate Demand for Insurance, Journal of Business 55, 281-296.
Mayshar, J., 1978, Investors' Time Horizon and the Inefficiency of Capital Markets, Quarterly Journal of Economics 92, 187-208.
McKenzie, R.B., 1986, Tax/Compensation Schemes: Misleading Advice in a Rent-Seeking Society, Public Choice, 48, 189-194.
Mishan, E.J., 1976, Cost-Benefit Analysis (Praeger Publishers, New York).
Newberry, D.M.G. and J.E. Stiglitz, 1982, Risk Aversion, Supply Response, and the Optimality of Random Prices: A Diagrammatic Analysis, Quarterly Journal of Economics, 97, 1-26.
Nisbett, R. and L. Ross, 1980, Human Inference: Strategies and Shortcomings of Social Judgment (Prentice-Hall, New Jersey).
Niskanen, W., 1971, Bureaucracy and Representative Government.
Olson, M., 1965, The Logic of Collective Action (Harvard University Press, Cambridge).
Pauly, M., 1968, The Economics of Moral Hazard: Comment, American Economic Review, 58, 531-536.
Pauly, M., 1974, Overinsurance and Public Provision of Insurance: The Roles of Moral Hazard and Adverse Selection, Quarterly Journal of Economics 88, 44-62.
Plotnick, R., 1981, A Measure of Horizontal Inequity, Review of Economics and Statistics 63, 283-288.
Plotnick, R., 1982, The Concept and Measurement of Horizontal Equity, Journal of Public Economics 17, 373-391.
Plotnick, R., 1984, A Comparison of Measures of Horizontal Inequity Using Alternative Measures of Well-Being, Institute for Research on Poverty Working Paper 752-84.
Poterba, J., 1984, Expected Future Tax Policy and Tax-Exempt Bond Yields, NBER Working Paper 1469.
Rosen, H.S., 1978, An Approach to the Study of Income, Utility, and Horizontal Equity, Quarterly Journal of Economics 92, 307-322.
Rothschild, M. and J. Stiglitz, 1976, Equilibrium in Competitive Insurance Markets: An Essay in the Economics of Imperfect Information, Quarterly Journal of Economics, 90, 629-49.
Shavell, S., 1979, Risk Sharing and Incentives in the Principal and Agent Relationship, Bell Journal of Economics 10, 55-73.
Shavell, S., 1980a, Damage Measures for Breach of Contract, Bell Journal of Economics 11, 466-490.
Shavell, S., 1980b, Strict Liability Versus Negligence, Journal of Legal Studies, 9, 1-25.
Shavell, S., 1981, A Note on Efficiency Versus Distributional Equity in Legal Rulemaking: Should Distributional Equity Matter Given Optimal Income Taxation?, American Economic Review, 71, 414-418.
Shavell, S., 1982, On Liability and Insurance, Bell Journal of Economics 13, 120-132.
Shavell, S., 1984, Liability for Harm versus Regulation for Safety, Journal of Legal Studies, 13, 357-74.
Skinner, J., 1986, The Welfare Cost of Uncertain Tax Policy, NBER Working Paper 1947.

Stiglitz, J., 1982, Utilitarianism and Horizontal Equity: The Case for Random Taxation, Journal of Public Economics, 18, 1-33.
Traynor, 1977, Quo Vadis, Prospective Overruling: A Question of Judicial Responsibility, Hastings Law Journal, 28, 533-xxx.
Watson, H., 1984, The Effects of Compensation for Uninsured But Insurable Losses, Journal of Risk and Insurance, 51, 498-512.
Weiss, L., 1976, The Desirability of Cheating Incentives and Randomness in the Optimal Income Tax, Journal of Political Economy, 84, 1343-52.
Wilson, C., 1977, A Model of Insurance Markets with Incomplete Information, Journal of Economic Theory, 16, 167-207.
Zodrow, G., 1981, Implementing Tax Reform, National Tax Journal 34, 401-418.
Zodrow, G., 1985, Optimal Tax Reform in the Presence of Adjustment Costs, Journal of Public Economics 27, 211-30.