Social Traps and Environmental Policy

Why do problems persist when there are technical solutions available?

Robert Costanza

In recent years scientific understanding and public awareness of environmental problems have increased tremendously. Yet our ability to manage these problems effectively has in too many cases gotten worse, not better. Consider the following examples:

- Thousands of inactive and abandoned hazardous waste dumps exist in the United States. There are thousands more active sites, plus deep well injection sites. These wastes are slowly seeping into drinking water supplies; many of their constituent chemicals have totally unknown effects on human and environmental health. Each year thousands of new chemicals are introduced into the environment whose effects are equally unknown. Superfund is only beginning to clean up the worst inactive and abandoned sites (most of the money so far has gone into planning and litigation). Handling active sites is largely an individual state’s responsibility, and resources for this critical task are too often insufficient.

- The coastal zone of Louisiana is eroding at a rate of more than 100 km²/yr, and the rate is accelerating (Gagliano et al. 1981). Canal dredging in the coastal zone has been shown to be a significant contributing factor to wetland erosion (along with sediment diversion resulting from leveeing the Mississippi River and natural subsidence) (Craig et al. 1979, Deegan et al. 1984, Leibowitz et al. 1987, Scaife et al. 1983). In 1980 the state implemented an office of coastal zone management to permit canal dredging and other activities in the coastal zone. Since 1980 there have been several thousand canal-dredging applications (mainly for the development of oil and gas), and the office has yet to deny a permit. Some small-scale projects designed to reverse wetland erosion have been implemented but have been largely ineffective. The state is currently considering several options to control the erosion of wetland, all of which will be very expensive.

- Probably the biggest environmental threat facing the world today is the threat of nuclear war. The likelihood that a nuclear winter will result from even a moderate thermonuclear exchange has been estimated by some researchers to be very high (Turco et al. 1983, White and London 1984). There is still considerable uncertainty about the global environmental effects of nuclear war. Even in the best-case scenario the damages would be astronomical, and in the worst case such an event could drive the human species to extinction. It would be catastrophic not only for human life, but for most forms of life on the planet (Ehrlich et al. 1983). There has, however, been little real movement toward arms control or other long-term solutions.

This litany could be continued indefinitely, spanning the spectrum from local to global problems (see also Borman and Likens, p. 370, and Catton, p. 413, this issue). The fundamental question is: Why do these problems persist in the face of adequate scientific understanding of their nature and of technical methods to solve them? What do they have in common? Can they be solved, and if so is there a general category of solutions?

I argue that all the environmental problems mentioned above (and many other social problems) belong to a category of phenomenon called social traps (Platt 1973). Like animal traps, social traps lead an unwary victim into the jaws of disaster with a tempting bit of bait, and, once the victim is caught, make escape extremely difficult. By studying the features real-world social traps have in common, and by experimenting with some simple laboratory examples of social traps, we can learn more about their general nature and the nature of effective escapes from them. A broad ecological perspective can be effective in understanding, avoiding, and escaping from some social traps, but it must be coupled with effective public
policy. Effective policy involves a range of activities from education to regulation to correcting the misleading short-term incentives (the bait) that create traps in the first place.

Social traps

A social trap is any situation in which the short-run, local reinforcements guiding individual behavior are inconsistent with the long-run, global best interest of the individual and society (Cross and Guyer 1980, Platt 1973, Teger 1980). We go through life making decisions about which path to take based largely on “road signs,” the short-run, local reinforcements that we perceive most directly. These short-run reinforcements can include monetary incentives, social acceptance or admonishment, and physical pleasure or pain. Figure 1a indicates the situation. In this figure the “correct” path (toward happiness and prosperity) seems obvious from the signs. In general, this strategy of following the road signs is quite effective in getting us where we want to go. Problems arise, however, when the road signs are inaccurate or misleading. In these cases we can be trapped into following a path that is ultimately detrimental because of our reliance on the road signs. For example, cigarette smoking is a social trap because by following the short-run road signs of the pleasure and social status associated with smoking, we embark on the road to an increased risk of earlier death from smoking-induced cancer. More important, once this road has been taken it is very difficult to change to another (as most people who have tried to quit smoking can attest).

Social traps can result from several causes. Table 1 is a taxonomy of these causes (Cross and Guyer 1980) along with some representative traps associated with each. Cigarette smoking, for example, is mainly a time-delay trap resulting from the fact that the positive and negative reinforcements are separated in time.

Traps can also arise out of simple ignorance of the relevant reinforcements, the change of reinforcements with time (sliding reinforcer traps), the externalization of some important reinforcements from the accounting system (externality traps), the actions of some individuals affecting the group in adverse ways (collective traps), or a combination of these causes (hybrid traps). These causes will be elaborated in the examples that follow.

Social traps are ubiquitous in everyday life and have become the basis for some important social psychology research (Brockner and Rubin 1985). For example, the “prisoner’s dilemma” game is a famous externality trap that has been used recently to study the conditions under which cooperation can evolve (Axelrod 1984). In this game two players must each choose either to cooperate or defect. If they both cooperate, they both reap a moderate reward (say three units each). If they both defect they both get a much smaller reward (say one unit each). If one cooperates while the other defects, the cooperator gets nothing (the “sucker’s payoff”) while the defector gets a reward larger than that for mutual cooperation (say five units). Under these conditions, if the players only meet once and cannot communicate with each other, the optimum strategy is to defect. For if one has no information about the other player, one must assume the worst to avoid the sucker’s payoff. The situation changes radically if the game is played several times with the same participants. Then each player has the record of their opponent’s past behavior, and the optimum strategy is not obvious.

Axelrod (1984) held a computer tournament pitting various submitted strategies against one another in a
round robin tournament of the iterated prisoner’s dilemma. He found that a simple strategy called tit for tat, which cooperates on the first move, then does whatever its opponent did last time, won the tournament. In the iterated prisoner’s dilemma, one can fall into the always-defect trap by not looking beyond the current move. If a strategy is too shortsighted it misses the opportunity to reap the benefits of mutual cooperation. Tit for tat was only a little more farsighted (one move) than all defect, but this was enough to dramatically improve its performance without making it too susceptible to being sucker-punched.

It turns out that there are several other strategies that would have won the tournament had they been entered. One of these cooperates on the first move and then looks at the entire past history of its opponent’s moves to generate a time-weighted expected value for the opponent’s next move. A strategy almost identical to this (called downing) was entered but lost because it defected on the first two moves. In the iterated prisoner’s dilemma, it pays to assume the best of your opponent (at least until you have been proven wrong) and to have a good memory.

The tragedy of the commons is another well-known social trap used to study overexploitation of natural resources (Hardin 1968). The classic commons trap goes something like this. There is a common property resource (say grazing land). Each individual user (rancher) sees his individual cost for consuming an additional unit of the resource (adding one more animal) as small and constant, and much less than the private benefits (from selling an animal). However, the overall cost to all the users of each additional resource unit consumed (animal added) increases exponentially as the resource is stressed. Eventually, one additional animal (which costs its owner no more than the first) leads to the destruction of the resource (which costs the animal’s owner and the rest of the ranchers tremendously). The tragedy of the commons is a collective trap that occurs because the costs and benefits apparent to the individual are inconsistent with the costs and benefits to the collective society.

Edney and Harper (1978) experimented with a simple game designed to test people’s behavior in a commons game. In this game a pool of resources is represented by poker chips. The resource pool is renewable; it is replenished after each round in proportion to the number of chips left in the common pool. The objective for each player is to accumulate as many chips as possible from the common pool. At each round, players can take either one, two, or three chips. If all players take three chips per round, the resource pool is quickly depleted, and the players end up with far fewer chips than if they had all taken only one chip per round, since doing so would have allowed the resource pool to replenish itself. This game is a trap (and a good analogy for many real-world commons-property resource problems) because the short-term, narrow incentives (to take as many chips as possible each round) are inconsistent with the long-term incentives (to accumulate as many chips as possible by the end of the game).

### Escaping social traps

Cross and Guyer (1980) list four broad methods by which traps can be avoided or escaped. These are education (about the long-term, distributed impacts); insurance; superordinate authority (i.e., legal systems, government, and religion); and converting the trap to a trade-off (i.e., correcting the road signs). Figures 1b–1d illustrate these methods using the road sign analogy.

Education can be used to warn people of long-term impacts that cannot be seen from the road (Figure 1b). Examples are the warning labels now required on cigarette packages and the warnings of environmentalists about future hazardous waste problems. People can ignore warnings, however, particularly if the path seems otherwise enticing (e.g., warning labels on cigarette packages have had little effect on the number of smokers).

The main problem with education as a general method of avoiding and escaping traps is that it requires a significant time commitment on the part of individuals to learn the details of each situation. Our current society is so large and complex that we cannot expect even environmental professionals, much less the entire voting public, to know the details of all the extant environmental traps. For education to be effective in avoiding traps involving many individuals, all the participants must be educated. For example, in the renewable resource game mentioned earlier if any one of the players starts taking three chips per round the others must follow suit. If any one player consumes three chips per round, the other players will do individually worse by restricting their consumption to one per round than if they consume three per round. The trap can only be avoided if all the players restrict their consumption to one chip per round, which requires that they all are educated about the nature of the trap.

The superordinate authority approach to avoiding social traps is illustrated in Figure 1c. Governments can forbid or regulate certain actions that have been deemed socially inappropriate. The problem with this approach is that it must be rigidly monitored and enforced, and the strong short-term incentive for individuals to try to ignore or avoid the regulations remains. A police force and legal system are very expensive to maintain, and increasing their chances of catching violators increases their costs exponentially (both the costs of maintaining a larger, better-equipped force and the cost of the loss of individual privacy).

Religion can be seen as a much less

---

**Table 1.** A short taxonomy of social traps (Cross and Guyer 1980).

<table>
<thead>
<tr>
<th>Cause of trap</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time delay</td>
<td>Discounting, smoking, drug addiction in general</td>
</tr>
<tr>
<td>Ignorance</td>
<td>Slot machines, gambler’s fallacy</td>
</tr>
<tr>
<td>Sliding reinforcer</td>
<td>Pesticide overuse</td>
</tr>
<tr>
<td>Externality</td>
<td>Pollution, prisoner’s dilemma</td>
</tr>
<tr>
<td>Collective</td>
<td>Tragedy of the commons</td>
</tr>
<tr>
<td>Hybrid</td>
<td></td>
</tr>
</tbody>
</table>
expensive way to avoid certain social traps. If a moral code of action and belief in an ultimate payment for transgressions can be deeply instilled in a person, the probability of that person’s falling into the “sins” (traps) covered by the code will be greatly reduced, and with very little enforcement cost. On the other hand, there are problems with religion as a means to avoid social traps. The moral code must be relatively static to allow beliefs learned early in life to remain in force later. And it requires a relatively homogeneous community of like-minded believers for religion to be truly effective. This system works well in culturally homogeneous societies that are changing very slowly. In modern, heterogeneous, rapidly changing societies, religion cannot handle all the newly evolving situations, nor the conflict between radically different cultures and belief systems.

A trap may be changed to a trade-off by imposing compensatory fees (Figure 1d). Many theorists believe that this method is the most effective in avoiding and escaping from social traps because it does not run counter to our normal tendency to follow the road signs; it merely corrects the signs’ inaccuracies by adding compensatory positive or negative reinforcements.

A simple example illustrates how effective this method can be. Playing slot machines is a social trap because the long-term costs and benefits are inconsistent with the short-term costs and benefits (Cross and Geyer 1980). People play the machines because they expect a large short-term jackpot, while the machines are in fact programmed to pay off, say, $0.80 on the dollar in the long term. People may “win” hundreds of dollars playing the slots in the short run, but if they play long enough they will certainly lose $0.20 for every dollar played. To change this trap to a trade-off, one could simply reprogram the machines so that every time a dollar was put in $0.80 would come out. This way the short-term reinforcements ($0.80 on the dollar) are made consistent with the long-term reinforcements ($0.80 on the dollar), and only the dedicated aficionados of spinning wheels with fruit painted on them would continue to play.

In terms of Edney and Harper’s common property resource consumption game, one could turn the trap into a trade-off by taxing any consumption above the optimum level for resource stability. For example, if players took two or three chips they could be taxed one or two chips respectively, so that the short-term benefits of taking more than one chip were offset by short-term costs. This would remove the short-term incentive to take more than one chip and make the long- and short-term incentives in the game consistent, thereby eliminating the trap.

Traps and environmental policy

Our approach in the United States to avoiding and escaping environmental traps has been mainly education and governmental regulation (Figures 1b and 1c). Although these methods are essential elements in the overall picture, they may not be the most effective means available in some situations. Converting traps to trade-offs seems to be a more effective method in many experimental trap situations, but it has been little used in the environmental area. In terms of environmental management, converting traps to trade-offs implies determining the long-run, distributed costs of environmentally hazardous activities and charging those costs to the responsible parties in the short run. Pollution taxes are the best-known example of this approach (and they have been quite effective in the few cases in which they have been tried), but it is possible to extend it to a much broader range of environmental problems. Below I give examples of how this approach might be applied to three critical current problems: hazardous waste management, coastal wetland management, and the nuclear arms race. I do not imply that this approach should be used independently of education, insurance, and regulation, but as an adjunct that can make the entire system more effective.

Hazardous waste management

The hazardous waste management problem can be viewed as a hybrid trap containing elements of time-de-
lay, ignorance, externality, and collective traps. The negative environmental effects of hazardous waste do not become evident until long after they are produced (time delay); their ultimate effects are largely unknown at the time of their production and release into the environment (ignorance); the negative effects are borne by parties other than the producer without sufficient compensation (externality); and common property resources are consumed (i.e., groundwater contamination) by individual agents who do not bear the costs of that consumption. The current regulatory approaches to hazardous waste management will not be discussed here.

To turn this trap into a trade-off we must charge the producers of hazardous waste for the ultimate long-run environmental and health costs of these wastes, and the charges must be imposed at the time of the waste’s production. Superfund can be seen as a small step in the right direction, but it relates only to the cleanup of abandoned hazardous waste sites, is assessed long after the wastes are produced, and is severely underfunded for the task.

The ignorance component of the hazardous waste management trap is the most difficult to deal with. How can we charge producers of hazardous wastes for the ultimate long-run costs of their waste if we have no idea what (if any) those costs will actually be? Part of the problem is that this ignorance and uncertainty about future costs is itself a cost, or more precisely a risk of unknown magnitude. The trap exists in large part because the producers of the hazardous waste do not bear this risk in the short-run, but pass it on to the general population and future generations.

One way to solve this problem involves changing our concepts of innocence and guilt as they apply to industrial products. As in our criminal justice system, there is currently a tacit presumption of “innocent until proven guilty” for chronic effects, particularly those that are difficult to quantify. The effect of smoking on lung cancer is a case in point; cigarettes had been presumed innocent until the overwhelming weight of evidence proved them guilty. Although such presumption of innocence is to-
tally appropriate for persons charged with crimes, it is not appropriate for hazardous or potentially hazardous materials, which we should instead presume guilty until proven innocent.

Thus, to turn the hazardous waste trap into a trade-off, we could charge producers of potentially harmful substances the worst-case costs. By worst-case we mean our current best estimate of the largest potential damages, although future information may later be used to increase the worst-case costs. This money could be put into a trust fund that would be returned to the producer (with interest) if and when the waste is proved to be innocuous, or could be used to compensate for damages caused by the waste if and when they become apparent. This procedure would change the short-term incentive structure from one that discourages waste producers from doing research on harmful effects, to one that encourages these studies so that some of the trust fund monies could be returned. The incentives to eliminate ignorance about hazardous waste effects would increase. It would also discourage production of large quantities of chemicals whose negative effects are unknown by making them as expensive in the short run as their risk cost to society.

Coastal wetland management

The problem of coastal wetland management is another example of a complex hybrid trap, with time-delay, ignorance, and collective elements. Coastal erosion in Louisiana is a particularly severe example. Canal dredging and other hydrologically disruptive activities have contributed to a current land loss rate of over 100 km²/yr (Craig et al. 1979, Gagliano et al. 1981, Scaife et al. 1983). It may already be too late to arrest or reverse this trend. The situation is a trap because the narrow, short-term incentives of those damaging the wetlands are inconsistent with the long-term good of the system.

To turn this trap into a trade-off, one should charge the responsible parties, at the time the damage-causing action is taken, the full cost of the ultimate environmental damage. To do this one needs to know the economic value to society of coastal marshes and the amount of marsh destroyed by each activity. As with the hazardous waste issue, there is much uncertainty involved in these estimates, but the worst-case costs should be assumed, and the burden of proof that the damages are in fact less than the worst case should be shifted to the parties who caused the damage. In addition, there needs to be a system for reducing the costs in proportion to any reduction in damages below the worst case.

For example, a recent study concluded that each acre of coastal wetlands in Louisiana has a present value to society of approximately $2500–$10,000/acre (Farber and Costanza 1987). Increasing the accuracy of the valuation estimates is an expensive proposition, and one that would stress the state's research budget.

To effectively eliminate this trap, one could charge the parties responsible for marsh destruction (i.e., oil companies for dredging access canals through wetlands) the $10,000/acre worst-case cost. These fees would go into a trust fund to be used for mitigating environmental damages by purchasing marshland elsewhere, backfilling canals, and diverting sediments. The responsible parties could lower the fee by proving that the damages are actually less than the worst-case assumption (by funding independent studies) or by minimizing the amount of wetlands they damage in the process of accomplishing their goal (e.g., by directional drilling or immediate backfilling). In either case the cause of wetland conservation would be served without unduly hindering the search for oil and gas.

The arms race and the dollar auction game

The current arms buildup is an environmental issue both because of its ongoing use and its potential for massive environmental destruction in the event of a major war (Mosley 1985, Sivard 1984, Wallace et al. 1986). Gradual deposal of the environment by more conventional means is certainly preferable to destroying it all at once in a nuclear war, but both must be included in any litany of major environmental problems.

The arms race has been escalating for the last 40 years and threatens to continue on this course with new developments, such as the Strategic Defense Initiative. It can be argued that the costs and risks of the arms race far outweigh its benefits, yet the process continues with no sign of abatement.

One simple but enlightening model for the study of the escalation process is known as the dollar auction game. This game is a social trap that was designed specifically to simulate conflict escalation (Shubik 1971). The dollar auction is just like a normal auction except that both the highest and the second-highest bidder have to pay the auctioneer their bid at the end of the game, but only the highest bidder gets the prize. This rule leads to some unexpected behavior that is a useful model of the arms race. Players in the dollar auction game frequently bid much more than $1 for a $1 prize—an irrational result that is the product of a series of "rational" decisions by the bidders, because the structure of reinforcements in this game is a trap. If player A had bid $1 and player B had the second-highest bid at $0.95, player B reasons that if he drops out he loses $0.95 while if he raises to $1.05 he only loses $0.05 (assuming he wins the $1 prize). So he usually raises, and this pattern of "rational" escalation (beyond the point where the overall outcome is rational) continues quite often to well beyond the $1 point. Individual and group behavior in the dollar auction game has been extensively studied by Teger (1980).

The dollar auction game can be converted to a trade-off by adding a "bidding tax" large enough to make dropping out rational in both the short run and the long run (Costanza 1984, 1986). For example, if when player B was at $0.95 he was told that it would now cost $2 to enter a bid of $1.05 (a $0.95 bidding tax), he would reason that if he drops out he loses $0.95, but if he raises he loses $1 even if he wins the prize! So the chances that he would drop out and escape the trap are increased. This method has proven to be effective in experiments using the dollar auction game (Costanza 1986). Extrapolating these ideas to controlling the escalation of the arms race is tricky, but it is possible that a nuclear weapons tax (which
has already been proposed as a method to fund Third World development) may prove to be an effective control method (Costanza 1984, 1986, UN 1982).

**Conclusions**

Social traps are one way of generating those situations in which the short-term, local optimizing of individuals goes afoul. In this sense they indicate imperfections in the free market approach to resource allocation, which relies on short-term, local optimizing of individuals. It can be argued that the proper role of a democratic government is to eliminate social traps (no more and no less) while maintaining as much individual freedom as possible. This can be accomplished most effectively by turning the traps into trade-offs that can be handled within the current market system as modifications to the cost of potentially entrapping activities.

Social traps abound in environmental issues because of the abundance of imperfectly owned and common property resources. To turn these traps into trade-offs, we must calculate the long-term social cost of activities with environmental impacts and charge those costs to the responsible parties in the short run. If there are any uncertainties about the magnitude of these social costs, the worst case should be assumed, and the burden of proving that costs are actually less than this should fall on the parties responsible for the impact, not the general public. It is also critical that short-term incentives for reducing environmental costs (by reducing the charges) be included. We must make protecting the environment as economically attractive to individuals in the short run as it is to society in the long run.

All animals capable of choice can be trapped with the right bait. Intelligent ones can learn to avoid traps. Ingenious ones can even escape from traps. If we are to survive, we need to exercise our vaunted intelligence and ingenuity to see, avoid, and escape from the many complex traps we have laid for ourselves.

**Acknowledgments**

This research was supported in part by a Kellogg National Fellowship. J. Bartholomew, H. E. Daly, B. Han- non, and G. Suter provided useful comments on earlier drafts of this article. Special thanks to E. Chu for her help. This paper evolved from a presentation at the 1985 AIBS annual meeting in Gainesville, FL.

**References cited**


