

# Social Traps

JOHN PLATT *Mental Health Research Institute  
University of Michigan*<sup>1</sup>

A new area of study is the field that some of us are beginning to call *social traps*. The term refers to situations in society that contain traps formally like a fish trap, where men or organizations or whole societies get themselves started in some direction or some set of relationships that later prove to be unpleasant or lethal and that they see no easy way to back out of or to avoid.

Two recent descriptions of traps of this kind have already become widely quoted and discussed. The first is Garrett Hardin's (1968) article entitled "The Tragedy of the Commons." The title refers to situations like that of the Commons, or public grassland, of the old New England villages, where anyone could graze his cows freely. Since this is a "free good" for the owners of cows, every owner can make money faster by increasing the number of cattle that he grazes there. But as everyone's number of cattle increases, the grass gets scarcer until finally it is destroyed entirely, and the owners collectively wind up with a loss rather than a gain. The trap is that each individual owner continues to do something for his individual advantage that collectively is damaging to the group as a whole.

Hardin saw this as the prototype and formal analogue of the world population problem, where each family may find pleasure and advantage in more babies; and the problem of competitive consumption of nonrenewable natural resources; and

the problem of competitive extermination of the last great whales.

A converse type of situation might still be regarded as a generalized trap, but perhaps is more accurately called a *countertrap*. The consideration of individual advantage prevents us from doing something that might nevertheless be of great benefit to the group as a whole. It is, so to speak, a *social fence* rather than a social trap.

A famous, or infamous, example of this kind was the Kitty Genovese murder in New York City a few years ago, in which a girl was raped and killed in an areaway while more than 30 neighbors watched out the windows—and none of them called the police. This apparent failure of concern and action produced a national wave of horror, as well as much recrimination afterward among those involved. Yet, in such a situation, it is clear that there is a certain individual barrier against calling the police. Not only must you tear yourself away from the spectacle, but you face the probability of having to testify in court and even a chance of being hunted down by the murderer or his friends. Each observer may have felt a strong prick of social conscience at the time, but simply hoped that someone else would make the troublesome phone call first.

Many contrasting cases of this kind have been discussed in a fascinating article by Thomas Schelling (1971), "The Ecology of Micromotives." Schelling cataloged several dozen type situations where individual actions or inactions controlled by immediate personal goals or self-interest, even rather weak self-interest, produce long-range societal effects which are to almost no one's self-interest.

For example, he demonstrated how a population of red people and green people distributed at random over a chessboard who move from time to time to new sites can become sharply segregated by color very quickly even if the individuals have

<sup>1</sup>This article is a report of research in progress by John Cross, Mel Guyer, Gardner Quarton, and the author, all of whom are affiliated with the Mental Health Research Institute of the University of Michigan. The article was an invited address presented at the annual meeting of the American Psychological Association, Honolulu, Hawaii, September 1972.

Requests for reprints should be sent to John Platt, who is on leave for the 1972-1973 year. His address is Center for Advanced Study in the Behavioral Sciences, 202 Junipero Serra Boulevard, Stanford, California 94305.

only a mild preference for neighborhoods of the same color.

Another example is the decay of railroad service, as people begin to prefer using their cars. As the railroad begins to go downhill, still more people prefer cars. The process is self-accelerating, ending up with no one riding the trains, while there are traffic jams on the highways in which everyone involved would prefer—too late—to be using the railroad. The process of inflation is likewise self-accelerating, with every increase in inflation producing new demands for raises in wages and profits, which drive up inflation faster. When we begin to look at such examples, we see that many of our really troublesome social and political problems today are made difficult, not by stupidity or avarice or immorality but by a certain trap component of this kind.

Our group at the University of Michigan became interested in these questions in the course of studying Skinnerian mechanisms of reinforcement of behavior, examining how they apply to personal self-control and to social transactions. We were trying to make formulations of the behavioral results that might be applicable to several disciplines, my own interest being that of general systems theory; John Cross, an economist, was concerned with microeconomics and bargaining; Mel Guyer, a game theorist, was interested in the locked-in conflict and cooperation modes of non-zero-sum games; and Gardner Quarton, a psychiatrist, was interested in the explanatory and therapeutic possibilities of behavioral reinforcement.

After reading the Hardin article and later the Schelling article, we suddenly saw that a number of their social trap and countertrap situations could be formalized in a reinforcement language. This quickly led to a useful classification of different kinds of traps, with interesting parallels between what had seemed to be very different problems. This formulation led in turn to several suggestions of personal and social methods of self-control for getting out of these traps. Here I want to outline some of these new findings.

### Reinforcements and Behavior

First, however, it may be helpful to describe our way of formalizing the Skinnerian results showing the effects of reinforcement on behavior. Skinner uses a three-term formulation, with the experi-

menter or the environment creating a situation or stimulus, S, in which an organism or subject emits some behavior, B, which is followed by some reinforcement or result, R. We find it helpful to write this S-B-R formulation on two lines, as follows:

B            B            B  
S R... S R... S R... ..,

where the top line represents the actions or outputs of the organism and the bottom line represents the actions or inputs of the environment, with an ongoing transactional relation between organism and environment. The S-B-R sequence, even if written only once, is thought of as being repeated over and over again in learning or maintaining behavior. A positive reinforcement or result,  $R^+$ , is defined as an environmental consequence that makes an initial behavior, B, more probable when that particular type of S occurs again. An aversive consequence or a negative result,  $R^-$ , is one that makes a given B less probable the next time S occurs.

This is equivalent to a feedback formulation, in which S is an input field to the organism, B is the motor output from the organism back to the environment, and R is the "reafferent stimulation," or change of input field from the environment which gives the changed "error signal" to the organism and closes the feedback loop, as seen in Figure 1.

A frequent objection to the Skinnerian formulation is that the definition of reinforcement or of a positive reinforcer is "circular," as in saying, "behavior is made more probable by positive reinforcement." But this is very similar to the useful Darwinian phrase, "the survival of the fittest," which

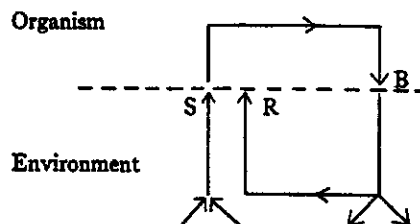


FIG. 1. Feedback loop where S = input field to the organism, B = motor output from the organism back to the environment, and R = the change of input field from the environment which gives the changed "error signal" to the organism and closes the feedback loop.

$$S = S_1 + S_2$$

$$S_1 = \text{physical/biological}$$

$$S_2 = \text{social, legal, inst'l}$$

$$R = R_1 + R_2$$

$$R_1 = \pm \text{Group}$$

$$R_2 = \pm \text{individ.}$$

$$S' \quad R' : S^2 \quad R^2$$

*politics*

*symptoms*

is likewise a circular definition of fitness. In fact, Skinner seems to think of the evolution of behavior in an organism from childhood as involving a similar "natural selection" of behaviors that "work" and therefore "survive" in a given stimulus situation, presumably because there are internal neurophysiological loops of response that are reinforced while others are eliminated. He sees the evolution of a repertoire of behaviors in an organism as like the evolution of species by survival in an ecosystem. "Reinforcers" are in fact defined much better than the Darwinian "fitness" of species, and reinforcers such as food, water, sex, petting, praise, and random jackpots of food or money operate similarly and predictably over a wide range of organisms.

What we realized when we began to consider our various social traps from this reinforcement point of view was that the trap depends on the difference between the personal or short-term reinforcements for a given B and the group consequences or long-term consequences of that B. Skinner has always emphasized that behavior is shaped more by its rapid consequences, within one second or a few seconds, than it is by what happens in five minutes or an hour—or at the end of the quarter when the student receives his grades. Immediate reinforcement singles out some particular recent behavior, while long-run reinforcement is ambiguous, not indicating which of thousands of previous behavioral acts is responsible for it.

In the *Federalist Papers*, Alexander Hamilton likewise emphasized as a central problem in designing a governmental structure the fact that men's behavior is more affected by immediate considerations of personal advantage than by the long-run public interest.

A social trap occurs, then, when there is an opposition between the highly motivating short-run reward or punishment,  $R_S^+$  or  $R_S^-$ , and the long-run consequences,  $R_L^+$  or  $R_L^-$ . In our notation, a trap then has the following form:

Trap:            B  
                  S  $R_S^+$  . .  $R_L^-$  .

Conversely, a countertrap or a fence would be written as follows:

Fence:            B  
(countertrap)    S  $R_S^-$  . .  $R_L^+$  .

Here, the immediate punishment (or its expectation after some experience) tends to block behavior B, even though there would be a long-run reward. In these formulations, we have a behavioral definition of the exact traditional meaning of the words trap and fence.

A similar concept of *reversal of reinforcers* can be applied to individual-group traps, where it is not a question of shorter and longer times so much as the fact that the personal reward or punishment,  $R_P^+$  or  $R_P^-$ , is in opposition to the collective or group advantage or disadvantage,  $R_G^+$  or  $R_G^-$ . Again we can have traps or fences depending on whether the initial personal result is positive or negative.

### Types of Traps

Using these ideas, our group has now studied some 40 cases and subcases and examples of various sorts, where the relation between  $R_S$  and  $R_L$ , or  $R_P$  and  $R_G$ , differs in one way or another. In this article, I discuss only the broadest general types. There seem to be three major classes: the *one-person* traps or self-traps; the group traps of the Kitty Genovese type or *missing-hero* type, where one person is needed to act for the group; and the group traps of the *Commons* type, where the common pursuit of individual goods leads to collective bads, because of scarcities, overcrowding, and the like. There can be both traps and countertraps in all three classes, although only a few of the possible subcategories will be illustrated here.

For clarity and simplicity, the different cases will be identified here by fairly abstract formulations and type anecdotes or mnemonic labels. However, I think in each case the reader will be able to see that these designate the trap aspect of several real social problems.

#### ONE-PERSON TRAPS

We thought it was important to study the various one-person traps first, to get their main features straightened out before going on to the group traps. The most important subgroup of one-person traps seems to involve the simple reversal of reinforcers after a *time delay*.

Such delayed reversals, where  $R_S^+$  changes to  $R_L^-$ , are exemplified in the smoking of cigarettes, where there is both a biochemical reinforcement and perhaps a social reinforcement in the short

run, but which may lead to lung cancer in the long run. Similarly with overeating, where there are the pleasures of the food and perhaps of a mother's approval in the short run, but an increased risk of heart attacks in the long run.

Countertraps of this type, with simple delayed reversals, where  $R_S^-$  changes to  $R_L^+$  in the long run, are exemplified by the difficulty of saving for Christmas, or for old age, because of the deprivation of present pleasures—even though the savings can eventually lead to a considerable reward, with interest.

A second subgroup of the one-person traps is that in which the problem is not simple delay, but rather *ignorance* of the unexpected or reversed outcome. The fish swimming into the fish trap does not know that he cannot get out. In the long run, ignorance is as lethal as evil. This is the case of the man handling a gun who shoots himself or his friend because he "didn't know it was loaded."

Another subgroup is that of *sliding reinforcers*. These are reinforcers that change steadily as you go on repeating a behavior, so that it becomes less and less rewarding and in fact punishing. Yet you may go on for a long time with the habit, or you may keep trying, in the hope that the results will sometime again be as good as they once were. This is one aspect of drug addiction, where the original kick and the fun you had with your friends turn into a frightful necessity which you regret for most of your waking hours.

The general public has a similar problem in the deterioration of old pleasures, such as the taste of food. Bacon these days has a label that reads "artificially smoked"; but it doesn't taste like smoked bacon to me, and I might never have gotten in the habit of breakfast bacon if it had tasted like that 50 years ago.

Today our global changes are confronting us with many sliding reinforcers. Once, large families with more babies were good for survival, and they were a delight, but now excessive babies have turned into an expense and have contributed to overcrowding for everyone. At one time, more consumption of natural resources and of electric power gave us consumer goods and liberation, but now we see them turning into a destruction of our natural heritage, with pollution and overheating.

#### THE MISSING HERO

When group profit,  $R_G^+$ , is blocked by  $R_P^-$  for any personal action, we have the missing-hero trap. A type case is the mattress problem, which is entertainingly described in Schelling's (1971) article. Consider the situation, on a summer Sunday evening, when thousands of cars are coming back from a Cape Cod weekend on a two-lane road and a mattress falls unnoticed from the top of a station wagon and lies in the northbound lane. All of the cars behind, being uncertain, go around the mattress, waiting for the cars in the southbound lane to go by, and the result is a traffic jam that backs up for miles.

Now who moves the mattress? The answer is, generally, no one. People far back in the line do not know what the trouble is and cannot help. And the drivers close to the mattress are thinking only of how to get around it quickly—and after they have spent so long in line they are damned if they will spend another several minutes, perhaps endangering themselves, to stop to move the thing. Those who have gone past, of course, no longer have any incentive for moving it.

In such a situation, it is true that sometimes a hero does come forward. Once, when I told an Englishman this story, he said, "Hah! That's only a problem for *Americans!* If there had been a single Englishman in that line, he would have gotten out and moved the mattress, because we are trained in childhood to take leadership in a case like that." This reminded me of another group that also would not have had such a problem—the Mormons. I was once at Utah State University in a snowy February, and we went to an undergraduate party in the mountains. The students, who were mostly Mormons, almost automatically formed a 14-car caravan up the icy winding road. They kept looking up and down the line to see if they were still together, and the whole caravan stopped several times with all the men getting out to push a car that had lost traction or was sliding off the road. So perhaps a Mormon would also have moved Schelling's mattress.

These examples immediately show the role played by moral or ethical training in preventing or getting out of this kind of group trap. Nevertheless, the willingness even of people of great goodwill to come forward and play the hero in such a case obviously depends a great deal on the

level of personal difficulty or danger. We see this in the reluctance of anyone, either in Sicily or America, to testify against the Mafia. The Kitty Genovese case, which belongs in this missing-hero category, may not indicate so much a lack of character in Americans as a different perception of personal hazard in "getting involved." We need continuing positive reinforcers for brave and intelligent initiative to help keep up our "character" in cases of this kind: combat ribbons and awards for valor on the civilian front, so to speak. (If it is not the Star of the Order of Lenin, perhaps it could be the Star of John Lindsay!)

### *Individual Goods and Collective Bads*

The third major category is that of purely collective traps, like the Tragedy of the Commons, where  $R_0^-$  follows only because of the excessive number of  $R_1^+$  practitioners. The problem cannot be solved by one or two heroes volunteering not to graze their cows on the commons, although such a course is frequently advocated by men of goodwill. And the problem is not the result of any single person doing anything that is unethical or bad, for if the number of persons involved were kept small, one can imagine that the collective good would be well served by the sum of all the personal  $R_1^+$  rewards. It is when the number is excessive that the difficulty arises.

A famous type problem which can be fitted into this classification is the Prisoner's Dilemma. This is one of the types of two-person non-zero-sum games which Anatol Rapoport (1966) has studied so extensively for many years. The type situation is that of two prisoners who have been caught by the police in some misdemeanor but who are suspected of worse crimes. They are held incommunicado from each other and each is questioned. The police offer a pattern of rewards such that if they both "talk" or "defect" on each other, they get the standard sentence for their crime; if they "cooperate" with each other, so that neither talks, they get off lightly for their misdemeanor; but if one talks and the other does not, the first gets a reward and goes free, while the second gets a doubly severe sentence.

In this situation, the *payoff matrix* is designed by the police so that each man benefits by defecting, no matter what his partner does. If his partner defects, the first gets only the standard

sentence, while if his partner makes the mistake of cooperating, the first man gets off with a reward. Yet if both cooperate with each other by not talking, they do much better than if they both defect. So individual rationality is at odds with collective rationality.

What do human beings do in such a situation? In our Institute, we have had thousands of such non-zero-sum games played by student volunteers for real money with various payoff matrices. Sometimes pairs or groups of students play against each other repeatedly for hundreds of trials, and sometimes they play without knowing it against a stooge or a computer which is programmed to respond in one regular pattern or another. Generally, in prisoner's dilemma situations, it is found that the opposing players tend to lock into either steady cooperation or steady conflict with each other. Which pattern is obtained seems to depend critically on the outcome—or should we say the "reinforcements"—of the first few plays. Sometimes a pattern of cooperation is quickly experienced as mutually profitable and is kept. But if such a pattern is not started early, it seems to be almost impossible for anyone to continue to cooperate when his opponent is continually defecting on him and making money at his expense. It is hard to keep working for  $R_0^+$  when the other party's behavior keeps turning it into  $R_1^-$  for you.

As Rapoport (1971) has emphasized, this dilemma and these alternative outcomes are remarkably parallel to some aspects of international relations in the non-zero-sum situations of either mutual economic dependence or mutual nuclear threat. The United States and Canada have had locked-in cooperation; the United States and Russia have had 25 years of locked-in hostility and arms escalation.

Another example of individual goods leading to collective bads, which can also be fitted mathematically into this same classification, is the Sell-A-Dollar game, which was invented by Martin Shubik a few years ago. This has some formal resemblance to the Prisoner's Dilemma, but it has an additional escalation feature. At a dull party, just to make things lively, I offer to auction off a dollar. You may laugh at this as absurd, for why should I auction it off for less than a dollar, and why should anyone pay more? But you agree to play just to see what will happen.

However, to keep the game interesting, I make some simple rules. The first is that bidding starts at a nickel. The second is that bidding must go up by 5¢ per bid. The third is that bidding must not go over \$50. (Everybody laughs.) The fourth and last rule is that since you will all try to get my dollar for so little, I want the two highest bidders both to pay me their bids, although only the highest bidder will get the dollar. (And *that* is where the kicker is.)

So the bidding starts—and when it gets up to about 25¢, you begin to wonder how high it will go. In a short time, you know. For there are two moments of truth in the game. The first one is when the bidding passes 50¢. Someone has bid 50¢, you have bid 45¢. And if you go on to bid 55¢, then this SOB with the dollar (me) will be getting back more than the dollar. But if you don't raise, you will lose 45¢. So you raise.

The second moment of truth is when the bidding passes \$1. The other fellow has bid a dollar, while you have bid 95¢. And if you raise to \$1.05, even if you win, you will lose money. But if you don't raise, you will lose a lot more. So you raise.

At this point, of course, I can lean back and watch you two clobber yourselves to death. The only limit on how high the bidding goes depends on how much money you have, or what your wife will think, or how furious you are with yourself. It is a dangerous game to play, because this series of escalating reinforcements and pressures can change the friendships or the relationships of everybody involved, perhaps permanently. I would not play it with children, even for pennies.

The only case I know where the bidding has been actually tested was when Layman Allen played Sell-A-Dollar with play money, with some of the people in his academic games groups. In one of his tests, the bidding for the dollar went up to \$4 before they tired of it.

Why describe this theoretical game in such detail? The reason is that it appears to represent in simplified monetary form some of the escalation aspects of such problems as drug addiction. The first little injection with your friends—the first bid—starts off easy and light, as a game, but the bidding then gets higher and higher until you are losing more than you are gaining. Yet with every shot you are getting again and again that

little  $R_P^+$  whose influence outweighs that great big long-run  $R_P^-$  you are steadily losing, as well as the  $R_G^-$  that your family and society are losing, too.

The Sell-A-Dollar game also throws further light on the escalation aspects of the international arms race. As Rapoport (1971) has pointed out in his recent book *The Big Two*, the military-industrial complexes of the United States and Russia are like two con men, who are actually working together in "selling a dollar" to our two governments—only the governments are raising their bids by \$10 billion each time, instead of a nickel. The immediate reinforcers of the bidding situation continue to be reinforcing to each group, thereby producing and maintaining behaviors of both countries which are extremely damaging to both of them in the long run, not only in expense and in terror, but finally, all too likely, in annihilation.

I hasten to say that there are ways you can prevent getting into, and can sometimes get out of, the sell-a-dollar traps. One first thing that can be done is to tell everybody about the game, so that people will be less likely to get into this kind of trap. A little preeducation and dramatic warning always help. In addition, the process of seeing the game as a whole—getting a metapicture, so to speak, of the competitive processes and the outcome—helps prevent a person from becoming quite so entrapped by the immediate reinforcers of each bidding step.

But what is most important is to see that it is possible to change the character of the game with side agreements and side payments. When the bidding passed 25¢, if you had said to your opponent, "You take the dollar and split the profit with me," you would both have made money, if he had had the sense to do so. The United States and Russia made a side agreement of this sort with the Atmospheric Test Ban Treaty, which has been to the advantage of all of us for 10 years now.

#### LOCKED-IN ASPECTS OF COLLECTIVE BEHAVIOR

It is worth digressing for a moment to note the locked-in behavior that is characteristic of many of these social traps. Immediate small reinforcements, or lack of them, lead to self-maintaining or stereotyped behavior in the mattress problem, in the conflict and escalation games, and in many other social situations.

Of themselves, of course, locked-in social behaviors and relations are not inherently either good or bad. A working society requires thousands of them to maintain the behavioral patterns that supply food or goods, just as a biological organism requires thousands of locked-in and repetitive enzyme cycles to maintain its metabolism. We are used to such contracts and networks, but what is surprising to us is the unexpected locked-in patterns that appear to arise from the free interactions of many independent individuals. And we still need to learn how to produce helpful ones and how to prevent or correct damaging ones in these collective relationships.

There are three distinct types of locked-in patterns in collective behavior that need much more study from a reinforcement point of view to see the microstructure that gives rise to a kind of social thermodynamics. The first type is what Adam Smith referred to as the *invisible hand* of the marketplace. He used this term to emphasize the absence of any overt or mechanical causal mechanism in the stabilization of prices or wages around some median value in a free economic market of competing individuals. A similar invisible hand tends to equalize and centralize the political parties in systems with majority (rather than proportional representation) elections.

The second type might be called by contrast the *invisible fist*, where the competition of numerous individuals does not produce agreement on a median value, but instead runs away from the median, with either escalation or elimination past some point of no return. This happens with Gresham's law in economics, where "bad money drives out good." Several of our current crises have this characteristic, as with the escalation of arms races or unrestrained pollution or the elimination of good railroad service, as we have noted. The urban crisis is made almost unsolvable by multiple complex escalations of this kind, as slum clearance drives slum dwellers elsewhere, and the poor are migrating into the city to get welfare money, while the rich are moving to the suburbs to escape taxes.

The third type of locked-in pattern could be called the *invisible chain*. This signifies a loop of transactional relationships among two or more people, forming self-maintaining systems that are sometimes very damaging and very hard to get out of. Married couples frequently get locked into repetitive disagreements over sex or money or the

temperature of the room or whether to go to the show early or late. Eric Berne (1964) has discussed various locked-in networks of this kind in his book *Games People Play*. In his game of "alcoholic," for example, he shows how the alcoholic is trapped in a self-maintaining game with three or four other people, such as the long-suffering wife, the best friend, and the corner bartender, with each of their responses reinforcing the others for *their* responses.

Over a lifetime, our originally accidental roles in many such chains, beneficial and damaging, may come to create and maintain the responses that finally appear as our personal or social "character." The self-maintaining character of federal bureaus or of the military-industrial complex come from large-scale invisible chains of the same sort. A careful analysis of such spontaneous lock-ins could be crucial today.

### Ways Out

It is to be emphasized that this type of formal analysis, classification, and explanation of many social problems as arising from reversal of reinforcers and the like is radically different from the usual explanations by moralists and social philosophers. For example, it is more immediate and practical than the explanations of some anthropologists and ethologists who imply that our social problems of conflict and disorganization are due to our evolutionary inheritance of "aggression" or of a "territorial imperative" (ignoring the fact that some societies have orders-of-magnitude of less aggressiveness or territorial demands than others). It passes by the explanations of therapists who interpret our personal and social problems as due to childhood frustrations or the Oedipus complex or to our cultural quest for power or the denial of love or the death wish. It is more behavioral and realistic than the countercultural claim that the locking in of our economic system to technology and consumerism and the increasing use of power and resources is due to Bacon or to "Newton's single vision" and the ignoring of Blake's fourfold vision.

Others have looked for explanations of social problems within the individual psyche. Koestler (1968) sees our problems today as basically due to a conflict between the lower instinctive brain and the recently evolved higher rational brain, and he

*invisible chains?*

has called for some drug which we could take to harmonize the two. Kenneth B. Clark (1971) has seriously recommended that antiaggressiveness drugs be given to presidents and public officials, although he did not discuss who will control the injections.

These two-level "mind-body" individual approaches to problems are like those of St. Paul. Like the other Judeo-Christian moralists, Paul interpreted lack of self-control and conflict and catastrophe as due to sin, or to the "old Adam" inside:

The good that I would, I do not; the evil that I would not, that I do . . . I see another law in my members, warring against the law of my mind. . . . Oh wretched man that I am, who shall deliver me from the body of this death? . . . with the mind I myself serve the law of God; but with the flesh the law of sin [Romans 7:19-25].

This mind-body dichotomy is what has led to the antisex attitudes and the asceticism and self-flagellation of Christianity, in its attempts to control and punish the lower members and drive out the devil. Niebuhr (1932) is not far from this same tradition in seeing social problems as being due to a kind of collective rather than individual wickedness.

But it is clear that Paul's "good" and "evil" could also be formalized as long-range or social consequences,  $R_L^-$  or  $R_L^+$ , of behavior which his "lower members" push him into doing or not doing because of the immediate gratification aspects of  $R_S^-$  or  $R_S^+$ . And it is clear that a change in the relationship between  $R_S$  and  $R_L$  to prevent this reversal of reinforcers can create easy self-control, or a society in which it is "easy to be good" without self-flagellation or repression. In spite of all of our serious problems and traps today, the poor are more or less fed, the children are taught, and the garbage is disposed of almost automatically by the reinforcements and feedback mechanisms of our society without the Christian effort and charity that were once necessary to solve such problems. We have learned to convert long-range social goods into daily wages,  $R_S^+$ , for social workers, teachers, and garbage collectors. Paul's view of the human condition turns to punishment and actually blocks this planned conversion of reinforcers that can give improved self-management or the correction of social traps.

648 • AUGUST 1973 • AMERICAN PSYCHOLOGIST

#### SOLUTION BY CHANGED REINFORCEMENT RELATIONS

In contrast to these usual prescriptions, the reversal-of-reinforcers approach suggests a number of specific changes of reinforcers or policies that can get us out of various social traps and that are already in effective use today for solving one problem or another. Some of these methods have been in use for hundreds or thousands of years, and much of our fashionable despair today comes from a kind of willful blindness to the methods that society developed long ago.

For example, the Tragedy of the Commons is essentially a problem of the allocation of scarce resources. And a half-hour's thought will turn up a dozen mechanisms that we use every day for dealing with such problems. In various societies, scarce resources of various kinds may be allocated by force, by tradition, by inheritance, or by election. When they are to be distributed to many people, they may be distributed by lot, or to the loudest voices, or by first-come-first-served, or by auction, or by selling tickets, as to a World Series baseball game. Fish and game commissions are set up, often with the support of the fishermen and hunters themselves, to sell licenses, set bag limits, and limit the hunting season so as to maintain the ongoing resource undiminished. Hardin (1968) made out his New England cattle owners to be a good deal more stupid than they actually were. The problem is not a problem of thoughtless competition, but rather the problem of setting up a superordinate authority to handle the reinforcement mechanisms—the tickets and bag limits—for getting out of these traps.

In fact, when we look at possible reinforcement changes, we can make a fairly exhaustive formalism of ways to prevent or get out of various social traps. Five major ways stand out immediately:

1. *Change the delay* to convert long-range consequences into more immediate ones. Or, as Skinner (1969) put it, "Bring the consequences to bear on behavior." This is part of what we do when we put warning labels on cigarette packages, or when, in "deconditioning" methods to stop smoking, we put unpleasant-smelling transparent tape around each cigarette.

On a larger scale, the highways of Indiana and Ohio were once jammed and ugly, and the problem seemed hopeless until some social entrepreneurs persuaded the legislatures to set up toll road corpora-

see p 644  
re role of  
ethics

tions, which sold bonds and paid the construction companies and the workers to build new highways. The short-range pay and return on investments,  $R_S^+$ , was a conversion of the eventual benefit,  $R_L^+$ , that would accrue to the state and all of the drivers—who were indeed glad in the end to pay the toll for *their* immediate pleasure,  $R_S^+$ , in driving on the improved highway. This method of solving the unsolvable highway problem was much easier and more successful than any direct attempt to change the old highways by the laborious methods of appeals to goodwill or unpaid community spirit or coercion, all of which are full of  $R_S^-$ . Obviously a superordinate authority or corporation was a useful intermediate step. The well-established investment mechanism is a powerfully effective device for getting over any short-run barriers,  $R_S^-$ , of habit or conflict or complexity because it brings the long-range benefits closer in the form of an  $R_S^+$ .

Education is another mechanism that gives an immediate payoff to the student—at least when it is good education—in the form of attention or grades or job-training pay or intellectual satisfaction,  $R_S^+$ , all to make the large long-run payoffs more immediately visible and effective. (This is separate from the question of education in general as a method of avoiding social traps. It is obviously a good and needed method for anticipating all our social problems, but our modern problems in a high-education society demonstrate clearly that education is not enough, unless it is combined with, or used to design, specific reinforcement mechanisms in each case.)

2. Add counterreinforcers, such as social incentives or punishments, to encourage or discourage behaviors by their immediate  $R_S^+$  or  $R_S^-$ . This is supposed to be the main function of punitive laws, but they obviously have little preventive effect in many areas, except in things like traffic control, where the probability of detection is high and payment is relatively quick. The incentives provided by administrative law and contract law are much more effective in large-scale problems (as might be expected from the Skinnerian effectiveness of positive reinforcement), and a combination of taxes and incentives for institutions and corporations is rapidly changing such problems as pollution, civil rights, and women's rights in the United States today.

3. Change the nature of the long-run consequence,  $R_L^-$ . One way to do this is by new inventions. Once upon a time it was a sin to make love to a girl you were not married to, and God would punish you, both with syphilis and with a baby that would kill her in childbirth. Edward Gibbon noted the injustice of God in not giving the Romans syphilis for their immoralities; and the Victorian novels as well as Hollywood movies until recently have had this theme of necessary punishment for sexual sins (though not other sins). But with the invention of penicillin to stop syphilis, with anti-septic methods to stop childbirth fever, and with easy contraceptives to prevent having a baby in the first place, suddenly it is no longer a sin to make love. State legislatures are decriminalizing extramarital intercourse, and even many religious leaders are now emphasizing the sacred value, rather than the sin, of lovemaking.

There are many large-scale social problems where improved design and planning is what can change the nature of the long-range consequences. Today, because of thought and design, social security is made a law; new cities are designed and built; an international monetary system is set up; and a hundred old problems are transformed.

4. Add  $R_S^-$  for competing behavior, which will not lead to the bad long-range consequences. Drink a diet cola with saccharine instead of fattening sugar; smoke a pipe instead of cigarettes. This is an essential component in the Skinnerian methods of self-control. To avoid the card games every night in the dormitory, give yourself goodies for studying instead: for example, some treat such as candy or a phone call to your girl friend only after so many pages of study, or a mark on your chart when your study alarm clock rings after every six hours of work, with 10 of these marks entitling you to have an afternoon in the woods or a dinner out in your favorite restaurant. If one reinforcer doesn't work, another one may, and so progress can indeed be made, until the larger reinforcements of grades or parental approval or the "natural reinforcers" of your own growing competence begin to maintain regular study habits.

Such methods have been used with considerable success to improve school performance and family relations of delinquent children in Tharp and Wetzel's (1969) study, *Behavioral Modification in the Natural Environment*. Larger social examples would include the revitalization of an ailing auto

Gov't / power

Knowledge

(punishment?)

inst. 1

technology.

assembly plant by new management methods (participatory reinforcements) (see Guest, 1962).

5. *Get outside help in changing the reinforcement patterns* of locked-in loops. This is the main new component of Tharp and Wetzel's (1969) methods. The delinquent child gets reinforced for his behavior by the attention he gets in being scolded, the excitement of being chased by the police, or the admiration of his friends; and the teacher, parents, police, friends, and the child are caught together in an invisible chain of self-maintaining reinforcement transactions. Tharp and Wetzel tried to find "mediators"—a teacher or adult friend who could see the child's daily behavior and give him immediate reinforcers, such as marks in a book for increased attention or reading, with so many marks entitling him to extra television viewing or horseback riding on weekends. As the child's behavior began to change within a few weeks, teacher, police, and parents changed their attitudes, and his friends began to admire him for different things (and were reinforced by the network for their change of values also). At this point, Tharp and Wetzel, as the outside "therapists," were able to withdraw because the system had been flipped to a new self-maintaining mode.

In many of our personal and social traps, we cannot easily see or change the locked-in reinforcer network ourselves, and a skilled outsider can help start the change to a new pattern. Our "lower members," for example, our taste buds, get immediate rewards from the cookies; and it takes our higher cortex to examine the situation and to put the cookies in the refrigerator instead of leaving them on the table. Likewise, it is easier to save for Christmas if the local bank, as the outside "therapist," has made a contract with you to deposit your Christmas savings every month and to write it in your book. Industrial safety laws, insurance, and social security deductions all represent outside agencies that we have called in to protect improvident man.

6. *Set up superordinate authority* to present entrapments, to allocate resources, to mediate conflicts, and to redirect immediate reinforcement patterns to more rewarding long-range goals. The organization of fish and game commissions against the exhaustion of the game, the Sherman Anti-Trust Act against the escalation of monopolies, and a sheriff system with mayors and courts in a western frontier town all represent something more than

just an outside therapist. They represent the democratic creation of new superordinate authority able to manage and correct social traps that were leading to collective bads.

This has happened over and over in human history, as in the creation of the European Common Market; the Special Drawing Rights, the new international money managed by an international commission; and the International Whaling Committee, even though it does not have any teeth in it at the moment capable of controlling the Russian and Japanese competition for the last remaining whales. It may be easier to set up superordinate authorities when there are many competitors than when there are only two or three, because the special pleading or self-interest of a strong individual can be more easily dealt with by the rest of the group when they are numerous. Nevertheless, the process is never very easy, and it would be important to make theoretical and historical studies to see for what kind of social traps superordinate authorities can and should be set up and how it can be done most easily and effectively.

### *Nested Traps*

Finally, it is important to note that there are mixed traps and, in particular, *nested traps* that are much harder to solve than any of the simple traps we have discussed so far. Traps of this kind include the locked-in violence of United States communications media, books, and drama; delinquent gang behavior; and drug and alcohol addiction.

In the United States media, the methods and habits of violence and violence as a community excitement are demonstrated daily and weekly many orders of magnitude more often than, say, human affection, or daily problem solving, or even (horrors!) sexual love. The media are in the invisible fist of the competition for more sales or higher audience ratings, so as to get more advertising or profits. (The invisible fist is proven by the self-accelerating elimination of hundreds of leading newspapers and magazines over the last 20 years.) This media violence locks in, in turn, to a multiplication of violent acts and violent individuals in the community. Headline reports of hijackings or any other special type of crime produce immediate imitations. In addition, it locks in the children and the older consumers to the self-maintaining idea that this is the only important

*competition?*

kind of news or drama, and conversely that the violence of real war is just another television spectacular. No cure of these nested lock-ins to violence in our society may be possible without a super-ordinate authority that can change all three of these self-maintaining loops simultaneously.

Gang behavior is also locked in at several levels. Each member of the gang is reinforced by his gang partners, as a subsociety, for slashing tires or taking drugs or escalating to more daring things. In addition, the gang entity is reinforced by the non-zero-sum conflict situation with other gangs or the excitement of avoiding the police. It is also often tacitly supported by an adult neighborhood subculture that protects the members and applauds their daring and sees them as expressing suppressed resentments against the larger society. And finally, the gang, with growing experience, makes contact with and comes to be supported by, and locked into, the larger criminal subculture, which is in turn serving needs and demands of the larger society. All of these loops would have to be interrupted and changed to produce any general change in the gang problem. It may not be possible, except through an ideological revolution, or a total change in city and neighborhood structures and legal-criminal relationships.

Likewise for drug and alcohol addiction. The alcoholic, for example, is locked in first by his own biochemical need for alcohol. He is also caught in the invisible chain of reinforcements with family and friends, as illustrated in *Games People Play* (Berne, 1964). In addition, he is an important cog in the network of the corner bar and the competitive liquor industry. And finally the liquor industry itself—producing a certain percentage of alcoholics—is serving a cultural need which our civilization can probably no longer do without, with our social contacts and meetings and most of our major business and government and military decisions being lubricated by alcohol. In this nest of

traps, the individual alcoholics come to be like traffic casualties, which can be reduced in number but not eliminated if our society is to continue to function in the only way it knows how.

But whatever solutions may eventually be found in these more complex nested cases, it is clear that the approach by analysis of reinforcements and reinforcement loops offers important new clarifying explanations and new tools for any amelioration that may be possible. Social traps are not the only kind of social problems, of course. For example, traffic accidents are not traps, nor are many fights and conflicts of interest, or business failures where there was an expectation of risk from the beginning. But the social traps represent all of our most intractable and large-scale urban, national, and international problems today. And it seems possible that the study of social traps from this reinforcement point of view may be opening the door on a whole new discipline that could do more than almost any other academic study to illuminate and solve these locked-in collective problems.

#### REFERENCES

- BERNE, E. *Games people play*. New York: Grove, 1964.
- CLARK, K. B. The pathos of power: A psychological perspective. *American Psychologist*, 1971, 26, 1047-1057.
- GUEST, R. H. *Organizational change: The effect of successful leadership*. Homewood, Ill.: Irwin, 1962.
- HARDIN, G. The tragedy of the commons. *Science*, 1968, 162, 1243-1248.
- KOESTLER, A. *The ghost in the machine*. New York: Macmillan, 1968.
- NIEBUHR, R. *Moral man and immoral society*. New York: Scribner's, 1932.
- RAPOPORT, A. *Two-person game theory: The essential ideas*. Ann Arbor: University of Michigan Press, 1966.
- RAPOPORT, A. *The big two, Soviet-American perceptions of foreign policy*. Indianapolis: Pegasus, 1971.
- SHELLING, T. The ecology of micromotives. *Public Interest*, 1971, No. 25, 61-98.
- SKINNER, B. F. *Contingencies of reinforcement: A theoretical analysis*. New York: Appleton, 1969.
- THARP, R., & WETZEL, R. *Behavior modification in the natural environment*. New York: Academic Press, 1969.

TV violence  
Robin Hood  
law  
only

Chain  
- 109  
- smof  
- transfer in  
- counter transfer

