ISSN 0022-3433 Journal of Peace Research, Vol. 21, No. 1, 1984

Review Essay: The Nuclear Arms Race and the Theory of Social Traps*

ROBERT COSTANZA
Center for Wetland Resources, Louisiana State University

Social traps are those all too common situations in which people behave contrary to their own self-interest while making what appear to them to be rational decisions. These situations may provide useful models for explaining (and analyzing possible escapes from) the nuclear arms race. The literature on social traps is reviewed, concentrating on three recent books: (1) *Social Traps* (1980) by J.G. Cross and M.J. Guyer, which provides a taxonomy of traps and escapes: (2) *Too Much Invested to Quit* (1980) by A.I. Teger, which is a detailed experimental study of the dollar auction game, an easily repeatable social trap designed to study conflict escalation; and (3) *The War Trap* (1981) by B. Bueno de Mesquita, which is only tangentially about social traps but provides supporting data and models on the history of international conflicts. Possible escapes from the nuclear arms trap are reviewed and the idea of a weapons tax to convert the trap into a trade-off is proposed and discussed.

1. *Introduction*

At the dawn of the nuclear age, Albert Einstein made what has turned out to be a chillingly accurate observation. The bomb changed everything, he noted, except our ways of thinking. Now, almost 40 years after the explosion of the first nuclear device, the world is still wrestling with the paradox that more nuclear weapons appear to buy both security and insecurity. One new way of thinking about a broad range of social problems has evolved recently that may help explain the nuclear dilemma. A body of research that has come to be called the theory of 'social traps', and a specific, easily repeatable, social trap called the 'dollar auction game' offers some new insights and perhaps some long-term solutions to the nuclear dilemma.

The theory was first elaborated by John Platt (1973) to explain the all too common social situations in which people behave contrary to their own self interest while making what appear to them to be perfectly rational decisions. Individual situations like this have been recognized for some time, but Platt and subsequent researchers have begun to look for the general principles underlying all traps, as a way to understand how to escape. As Kenneth Boulding (1977) has pointed out: 'The analysis of these processes of perverse dynamics is the key to successful intervention in human betterment' (p. 84). Phenomena as disparate as drug addiction, overuse of agricultural pesticides, resource depletion, and the escalation of conflicts can be gainfully analyzed from this perspective. A recent book by Cross & Guyer (1980), two of Platt's colleagues at the Mental Health Research Institute at the University of Michigan, provides a taxonomy of social traps and highlights the potential usefulness of this way of thinking about social problems.

All traps, social and otherwise, have some features in common that make this a useful analogy. An effective animal trap lures the victim with the illusion of an easy, low cost, meal, only to impose a very high cost once the animal has gone too far. Likewise, social

* This work was supported, in part, by a W.K. Kellogg National Fellowship. I thank J. Bartholomew, R. Bloom, C.J. Cleveland, H.E. Daly, B. Hannon, H.M. Kennedy, B. Reid, F.H. Sklar, R.E. Turner, and two anonymous reviewers for providing useful comments on earlier drafts.

traps work on the differences between the perceived or apparent costs and benefits to the individual participants and the real costs and benefits. As Cross and Guyer put it:

> all involve individuals who use reinforcements like road signs, traveling in the direction of rewards and avoiding the paths marked by punishments. Generally, this is a good way for us to get where we would like to go. Occasionally, however, these road signs lead to unfortunate destinations. These are our social traps. (p. 13)

Traps are entered into because of misperceived costs and benefits on the part of the trapee (poorly marked roads), and, once entered, pursuit of the obvious ('rational') course of action (continuing to follow the signs) only makes matters worse. This misperception can result from a number of causes that form the basis for Cross and Guyer's taxonomy: (1) 'time delay' traps occur when benefits and costs are misperceived because they are separated in time; (2) 'ignorance' traps occur through simple ignorance of real costs and benefits; (3) 'sliding reinforcer traps' occur when costs and benefits change gradually with time, but this change is not adequately perceived; (4) 'externality' traps occur when the actions of others change the cost and benefit framework (i.e. the 'prisoner's dilemma', cf. Brams 1975); (5) 'collective' traps occur when costs and benefits perceived by individuals are not equivalent to costs and benefits to the collective (the 'tragedy of the commons' (Hardin 1968) is a classic example); and finally (6) 'hybrid' traps occur from combinations of the previous causes.

Ultimately, the theory of social traps addresses the old controversy about the rationality of human behavior. If individual agents behave rationally, then one might expect the results of their collective behavior to be rational. If irrational results occur, it must therefore be because the individual agents are behaving irrationally. Social trap theory provides an alternative explanation that allows rational agents to produce irrational results if the situation is 'trap-like' in nature. These agents exhibit a limited degree of rationality, enough to get them into the trap, but not enough for them to see the trap and avoid it. In the language of social choice theory (Green & Laffont 1979) they are caught in an 'incentive incompatable situation', or in more colloquial terms, a 'vicious circle'.

2. *The dollar auction game*

A relevant example of this type of situation is an easily repeatable social trap called the 'dollar auction game'. The game was invented by Martin Shubik (1971) to study the escalation of conflicts. Teger (1980) has recently studied the behavior of individuals playing this game in a wide range of laboratory and classroom situations. The game is identical to a 'normal' auction except that *both* the highest bidder and the second highest bidder must pay the auctioneer their bid, while only the highest bidder receives the dollar prize. Almost invariably, the playing of this game leads to bids of well over a dollar for the dollar prize. This obviously irrational result comes about not because of irrational behavior on the part of the bidders, but because of the trap-like structure of the game.

In large groups, the bidding is usually entered by a number of individuals, but quickly reduces to two individuals who continue the bidding. There are two distinct turning points in the game. The first is at \$.50. Player A has just bid \$.50 and player B (who is the second highest bidder at, say, \$.45) must decide whether to raise or drop out. If he raises, the auctioneer will make money, but player B still has a chance to win the \$1. If he drops out, player B loses the \$.45 he has already bid. Player B rationally weights his expected costs (a sure \$.45 loss) and expected benefits (a net gain of \$.45 if his bid of \$.55 is successful in winning the \$1 prize) and decides to raise to \$.55. The second turning point is at \$1. Player A has just bid \$1 and player B must decide whether to raise or drop out with a current bid of \$.95. If he drops out he loses \$.95. If he raises to \$1.05 and wins the \$1 prize, he loses only \$.05. He does the rational thing and raises to \$1.05. From this point on the bidders have

forgotten their original motive for entering the game (to win money) and are now concerned with minimizing their losses and saving face by beating the other guy. The bidding usually continues to well above $1, until resources are almost exhausted or a fight breaks out.

A 'rational' decision rule for player B in the dollar auction would be:

$$\text{Raise if: } pV - B_t > B_{t-1} \quad (1)$$

or

$$\text{Raise if: } p > \frac{B_t - B_{t-1}}{V} \quad (2)$$

where B_t = bid of player B at time t. Must be larger than player A's bid which in turn must be larger than B's bid at time t-1

V = value of the prize

p = perceived probability of winning the prize

For example, at the dollar turning point, B_{t-1} = $.95, B_t = $1.05, and V = $1.00. The perceived probability of winning (p) need only be .1 to make raising an apparently rational decision. The effect of the bidding increment can be seen clearly in equation 2. Small increments lead to lower minimum values for p. Bidders will often lower their bidding increment when they become unsure of p.

Looking back at the dollar auction, it is obvious that the only truly rational thing to do is not enter the game in the first place because the game is a trap. Once the game is entered by at least two bidders, their fate is sealed if they behave rationally from that point on, and they do not form a coalition against the auctioneer. For example, if there were only two bidders and they understood the nature of the game, they could agree beforehand that one would bid $.05 and the other would not bid and share the $.95 net winnings. Of course, the auction format with many potential bidders tends to discourage this kind of cooperation. In experimental versions of the game with only two players, the degree of possible cooperation is usually controlled as part of the experiment.

The dollar auction game is a useful model of the conflict escalation process in general, and the nuclear arms race in particular. The nuclear arms 'auction' has come down to the two superpowers bidding for the 'prize' of global political and economic domination and military security, but the cost of the bidding is now no doubt well above the value of the prize and the 'irrationality' of this result is evident to everyone, including the bidders. The bidders (and the world audience with them) are trapped by the structure of the game, in spite of behaving in ways that to them appear to be the only rational ones.

While this model of the arms race is far from perfect, it captures some essential aspects and allows experimental manipulations that can provide insights into the entrapment process and possible 'escapes' from the trap. Teger (1980) has conducted extensive studies of variations of the dollar auction game in laboratory and classroom situations. The game can be seen as a specific example of a more general 'investment' trap. Bidding continues in the dollar auction because the second highest bidder always feels he has 'too much invested to quit' (which is the title of Teger's book). Many other closely related investment traps exist and have been studied by Rubin & Brockner (1975), Schroeder & Johnson (1982), Brockner et al. (1982), Brockner, Shaw & Rubin (1979), and others. In Cross and Guyer's taxonomy, investment traps are hybrids involving elements of time delay, ignorance, and sliding reinforcer traps.

Teger experimentally investigated why people enter the dollar auction, why and when they quit, the various stages of the escalation process, and the physiological and personality correlates of escalation. All of these studies contribute insight to the nuclear arms race. The following excerpt summarizes Teger's findings concerning the bidder's motivation and goals:

> Although the reasons which led the subject to make bids in the auction were many and varied, once

> they had gotten deeply involved (as the bidding approached $1.00) they were all faced with the same dilemma — quit and lose their investments, or continue to bid. Once that had occurred, their initial motivations changed and became much more uniform. The reasons for continuing the auction fell into several categories: showing that they could be the best, saving face by avoiding the stigma attached to quitting, inflicting punishment on the other person who was seen as responsible for the predicament in which they now found themselves, and attempting to regain the money which they had already invested in the auction (p. 16).

These motives certainly apply to the countries involved in the nuclear arms race. The sliding reinforcer aspect of this investment trap is also apparent. People enter the dollar auction thinking they will win money and end up trying to effect a face-saving exit. Similarly, the countries that have entered the nuclear arms race did so thinking they were buying security (military, political, and economic), but have ended up trying to minimize insecurity and anxiety without losing face. Part of the trap is the subtle way the payoffs change with increasing bids. Early in the game the quest for positive payoffs predominates. Later this changes to trying to minimize losses. Teger also found that when the bidding approached $1.00 the bidders felt that they were being forced by the other bidder to continue. They felt the other bidder must be crazy to continue, and maintained a very egocentric perspective that prevented them from realizing that identical forces were motivating their opponent. This view of the opposing bidder seems to be a characteristic of the relationship between the U.S. and the Soviets on defense issues as well.

Teger also found that the personalities of the bidders had little effect on the eventual outcome of the game. They did, however, affect the style of play:

> The population from which the groups were drawn did not affect the bidding. Undergraduates, graduates, and faculty played the game approximately the same way. In departmental colloquiums, the person to bid over $1.00 ranged from graduate students, to department chairmen, to experts on game theory. The only classes to produce a noticeably different style were the ROTC [Reserve Officer Training Corps] classes. Although the bidding pattern was similar to that of the other groups, the audience reaction was quite different from what was observed in the other classes. When the bidding approached and exceeded the $1.00 mark, the class cheered the bidders on. When the bidders appeared reluctant to continue the bidding, the class urged them not to quit and, instead, suggested additional bids that they might make. Also, the ROTC classes *requested* additional auctions, and continued to bid past the $1.00 mark for two or three successive auctions. It was obvious that, in the ROTC classes, conflict was nothing of which to be ashamed, but rather it was something to seek out and something at which to excel (p. 18).

It is interesting to note that the pattern of increasing U.S. defense spending has changed very little with changes in the political party of the Administration, but the style and rhetoric has changed dramatically.

3. *The war trap*

One might object that while the dollar auction game is an interesting bit of psychology, the real international situation is so much more complex that the game's usefulness in explaining international conflicts is minimal. Another objection could be that while it is one thing to trap ignorant, short-sighted individuals, it is quite another to trap sophisticated world powers. Just how good a model is the dollar auction game for international conflict situations?

There are, of course, some important differences in the details of the dollar auction vs. the real situation. For example, the nature and value of the prize are well known and precisely quantified in the dollar auction, while there is considerable uncertainty about what the prize is, to say nothing of its quantitative value, in the real situation. Even the cost of the bids (military expenditures, weapons characteristics, the chances of a nuclear exchange, etc.) are subject to much uncertainty in the real situation. From the perspective of trap theory, however, one would expect a situation with more uncertainty to be potentially more entrapping than one with less, since the chances to misperceive the relevant costs and benefits are greater.

The dollar auction analogy is useful because it shows that traps of this kind can occur even with full information, and they are exacerbated by the imperfect information and uncertainty associated with real situations. The analogy requires two assumptions about the behavior of countries: (1) they act as if to maximize their own 'perceived' self-interest when addressing defense issues; and (2) this perception is somewhat narrow, short-sighted, and imperfectly representative of the group, and can be inaccurate with respect to the country's real, long-run self-interest. Some evidence for the validity of these assumptions in the international conflict arena can be found in Bueno de Mesquita (1981). He constructed an expected utility theory of international conflict and tested the theory against the historical record of interstate threats, interventions, and wars over the last 162 years. The results of this commendable effort 'strongly support the proposition that positive expected utility is necessary — though not sufficient — for a leader to initiate a serious international dispute, including a war'.

Bueno de Mesquita points out several cases where the initiator's perceived (expected) utility accurately predicted the initiation of conflict, but the perception was later found to be inaccurate and the conflict was either aborted or lost. For example, in the Prussian-Swiss conflict over the Neuchatel in 1856, the French and British support of Switzerland was underestimated by Prussia in her initial expected utility calculations. Once the conflict started, the error became apparent and Prussia decided to back down.

What Bueno de Mesquita's analysis makes clear is that countries exhibit the same form of limited rationality in making national defense decisions which makes individuals susceptible to the dollar auction and other traps. Bueno de Mesquita seems to have been unaware of the work on the social psychology of traps mentioned earlier. The title of his book (*The War Trap*) appears to be only coincidentally indicative of the fact that his conclusions are consistent with social trap theory. His analysis provides the much needed empirical evidence at the international level which studies of the dollar auction game and other experimental traps cannot provide. Together, the work of Bueno de Mesquita and the social trap theorists provide the theoretical and empirical basis for a new understanding of armed conflict.

4. *Designing escapes*

This conception of the nuclear dilemma as a social trap provides some insights and the basis for some proposals for escaping the arms race. Along with their taxonomy of traps, Cross and Guyer also provide a taxonomy of possible escapes and their relative effectiveness in various situations. The escapes fall into four major categories: (1) change or supplement the reward-punishment structure of the situation so that the trap becomes a trade-off; (2) protect the victims of traps, i.e. social insurance programs; (3) rely on a superordinate authority to impose laws and punishments which are designed to prevent entering traps and make escape from already entered traps easier; and (4) educate individuals to see and avoid traps.

As regards the nuclear arms race, options 4 and 2 are not applicable since we are already in the trap and promised compensation of the victims of a nuclear war would be of little help. Option 3 is the most frequently mentioned. If a 'world government' with the power to enforce decisions could be brought into existence, it would certainly be capable of eliminating the war trap. The operational problems with implementing this solution, as evidenced by the history of the United Nations, seem to be enormous and possibly insurmountable, however.

The current approach to arms control — direct negotiation between the bidders — seems also to be ineffective. Arms treaties in themselves do nothing to change the underlying pattern of reinforcers that form the basis for the trap and therefore rely on trust that the other party will behave contrary to his/her short run benefit. This trust is simply not there between the superpowers, and would be very difficult to create.

Option 1, above, has been given very little attention in the international arena, even though, as Cross and Guyer point out:

> Certainly the most promising approach to any trap problem would be to present reinforcers in a manner that would direct behavior along rational paths by eliminating the biases introduced by time delay, ignorance, or even the fact that the consequences of one's behavior may fall on someone else. This procedure would convert a trap into a trade-off, presenting the individual with a set of reinforcers that occur in close proximity to the behavior in question and which closely match the actual reward and punishment patterns that underlie the situation. The trap then becomes a simple choice situation in which rational and learned behavior are coincident (p. 35).

Can the arms race trap be converted to a trade-off and could this actually produce a viable escape? The following preliminary proposal is one example of how this might be done.

5. *A global weapons tax system*

From the perspective of social trap theory, building weapons systems is entrapping when the apparent, short run costs (building the weapons) are not equivalent to the real, long run costs (using the weapons). Weapons systems are entrapping if their use can cause much more damage than it costs individual nations to produce and maintain them. Most modern weapons systems exhibit this characteristic, and nuclear weapons exhibit it to a staggering degree. To convert this trap to a trade-off, the apparent short run costs of building weapons systems must be made equivalent to their real, long run costs. One way to do this would be to impose a tax on new weapons systems equivalent to the difference between apparent and real costs.

Before addressing the problems of implementation and finding the proper tax schedule, consider how this system functions in the context of the dollar auction game. It would be equivalent to imposing a tax on bidding. Imposition of this added cost can change the payoff structure of the dollar auction sufficiently to make *not* raising look as rational to the bidder as it does to the informed observer. The rational decision rule changes from the one given by equation (1) to:

$$\text{Raise if: } pV - TB_t > B_{t-1} \tag{3}$$

or

$$\text{Raise if: } p > \frac{TB_t - B_{t-1}}{V} \tag{4}$$

where T = a tax multiplier = 1 + the tax rate

The tax multiplier raises the bidding increment and thus the probability of winning (p) that must be assumed to make raising the rational choice. There is a critical tax multiplier that yields a value of $p > 1$ in equation (4). Since the probability of winning cannot be greater than one, values of the tax multiplier greater than this value would make raising irrational in the short run as well as the long run.

For example, if B_{t-1} = \$.95, B_t = \$1.05, and V = \$1.00 the untaxed decision rule (equation 2) indicates that a .1 probability of winning the auction would make raising to \$1.05 the rational decision. This looks like a good bet and the bid is usually raised. If a tax multiplier of $T > 1.86$ in the taxed decision rule (equation 4) were imposed, however, it would require a greater than 1 probability of winning the auction to make raising the rational decision. I have tried this method of ending the dollar auction in classroom situations and found that it is a very effective way to stop the bidding at any level. It is even more effective if some additional stipulations are added about the disposition of the tax revenue to equalize the losses of both bidders at the end of the game.

The problems with employing some system like this to control real arms races are implementation and enforcement. In the dollar auction, the auctioneer represents a superordinate authority who could just as easily end the game outright as impose a bidding tax to end it. In the real situation it is just this superordinate authority which is missing, and

is unlikely to emerge given the power of the weapons which are themselves the bidding instruments. Can a tax system be implemented and enforced from 'the audience' rather than by the auctioneer? It is, in fact, possible to effect such a solution in the dollar auction. It is almost the reverse of the ROTC example. In the ROTC example, the cost of bidding is actually reduced by the audience, by providing positive social pressure and encouragement that functions like a negative tax or subsidy. In most auctions the audience is passive, but it is possible for them to exert negative social pressure thus increasing the cost of bidding (the dollar cost plus the social pressure cost), and decreasing the likelihood of bidding beyond the dollar point.

It might be possible for the global audience to organize and impose a tax on the superpower's weapons bidding without resorting to a superordinate authority, in a manner analogous to the way labor unions were able to effect change without superordinate authority. In fact, actions such as the peace movement in western Europe and the U.S. can be seen as attempts to raise the cost of nuclear weapons bidding. The often cited problem with this strategy is that it raises the cost to only one of the bidders. This seems to reduce the effectiveness of the tax system since, unless the cost to both bidders is raised equally, the incentive of the taxed bidder to 'keep the other guy from winning' can lead to his/her ascribing an ever higher value to the prize to keep the decision to raise 'rational'. Keeping the world 'safe for democracy (or communism)' suddenly becomes more valuable than simply keeping the world.

Imposition of a direct, monetary tax on weapons is one approach that could raise costs equally without inducing increased valuation of the prize. Implementation of a monetary tax scheme would, however, be a major obstacle. It would require a form of authority itself to implement, enforce and redistribute the tax. This authority might not need to be comprehensive or superordinate, however. A large subgroup of non-nuclear powers might be able to cooperate to implement the tax, obtaining the necessary leverage by forming an 'anti-nuclear cartel' and threatening to boycott trade (especially in natural resources) with the nuclear powers. If the revenues from the tax were distributed to third world countries, the scheme could also be viewed as a way to meet third world demands for wealth redistribution. Another possiblity might be for individual countries to appeal to existing international economic bodies (i.e. the International Monetary Fund, or the World Bank) for either removal of or compensation for the risk to their life and property posed by nuclear weapons. It is only fair that the true costs of the weapons be paid by their bearers and not innocent bystanders, but the bystanders to the nuclear auction can no longer afford to stand idly by. They must participate in helping the bidders (and themselves) to escape. A tax system would allow the weapons builders to continue to make their own decisions about weapons escalation, but removes the trap by converting it into a trade-off.

Obviously, these are only preliminary proposals intended to raise questions and possibilities. Much additional research would be required before concrete proposals could be advanced. I only wish to point out a direction that peace research might fruitfully take. There has been too much effort devoted to studying how the current game is played and how to play it better, and not enough to devising ways to change the underlying structure of the game, a game that — if left unchanged — could be the final one for *Homo sapiens*. The dollar auction game and the theory of social traps provide a formal structure to help analyze the potential effectiveness of various escapes from the arms race. Our species seems to have both a unique susceptibility to certain traps and a singular ability at trap recognition, avoidance, and escape. We can still escape the nuclear trap, but only if we first recognize that it is a trap and behave accordingly.

REFERENCES

Boulding, K.E. 1977. 'Twelve friendly quarrels with Johan Galtung', *Journal of Peace Research,* vol. 14, no. 1, pp. 75-86.

Brams, S.J. 1975. *Game Theory and Politics.* New York: The Free press.

Brechner, K.C. 1977. 'An experimental analysis of social traps', *Journal of Experimental Social Psychology, vol. 13, pp. 552-564.*

Brockner, J., J.Z. Rubin, J. Fine, T.P. Hamilton, B. Thomas, & B. Turetsky, 1982. 'Factors affecting entrapment in escalating conflicts: the importance of timing', *Journal of Research in Personality,* vol. 16, pp. 247-266.

Brockner, J., M.C. Shaw & J.Z. Rubin, 1979. 'Factors affecting withdrawal from an escalating conflict: quitting before its too late', *Journal of Experimental Social Psychology,* vol. 15, pp. 492-503.

Bueno de Mesquita, B. 1981. *The War Trap.* New Haven, CN: Yale University Press.

Cross, J.G. & M.J. Guyer, 1980. *Social Traps.* Ann Arbor: University of Michigan Press.

Edney, J.J. & C. Harper, 1978. 'The effects of information, in a resource management problem: a social trap analog', *Human Ecology,* vol. 6, pp. 387-395.

Green, J. & J.J. Laffont, 1979. *Incentives in Public Decision Making.* New York: North-Holland.

Hardin, G. 1968. 'The tragedy of the commons', *Science* vol. 162, pp. 1243-1248.

Platt, J. 1973. 'Social traps', *American Psychologist,* vol. 28, pp. 642-651.

Rubin, J.Z. & J. Brockner, 1975. 'Factors affecting entrapment in waiting situations: the Rosencrantz and Guildenstern effect', *Journal of Personality and Social Psychology,* vol. 31, pp. 1054-1063.

Rubin, J.Z. & B.R. Brown, 1975. *The Social Psychology of Bargaining and Negotiation.* New York: Academic Press.

Schroeder, D.A. & D.E. Johnson, 1982. 'Utilization of information in a social trap situation', *Psychological Reports,* vol. 50, pp. 107-113.

Shubik, M. 1971. 'The dollar auction game: a paradox in noncooperative behavior and escalation', *Journal of Conflict Resolution,* vol. 15, pp. 109-111.

Teger, A.I. 1980. *Too Much Invested to Quit.* New York: Pergamon.