

Formal Games as Probing Tools for Investigating Behavior Motivated by Trust and Suspicion

ANATOL RAPOPORT

Mental Health Research Institute, University of Michigan

Every technology derives support from four sources: first, a group of sciences which nurture its inventions; second, a group of techniques, the arts of the technology; third, a social organization which makes possible the implementation of the technology; and finally, an ethos which justifies its practices and its aims.

To take some obvious examples, modern industrial technology derives from physical science, from techniques of mass production, from certain forms of social organization (e.g., capitalist or socialist), and from an ethos which declares that affluence is preferable to poverty and that man is privileged to modify nature to suit himself.

Modern medical technology derives from physical and biological sciences, from certain organizational techniques (e.g., management of hospitals), from the same forms of social organization, and from an ethos which declares that human life is to be prolonged under almost any circumstances.

Military technology derives from similar scientific and social sources and from an ethos which declares that people may be killed on a mass scale in order to enhance the power and prestige of national states or in order to preserve forms of social organization.

The development of a technology obviously depends on the ways its various

sources of support interact with each other and with it. In all the examples so far mentioned, a sort of positive feedback has been operating to enhance an accelerating growth of each. It happens at times, however, that the interaction is of an inhibitive kind which ultimately destroys the technology. An example, which anthropologists are fond of quoting, is that of a certain South Pacific people who had developed an excellent canoe-building technology but along with it a set of rituals so complex that in time no canoe could be built without violating one of innumerable taboos with its concomitant danger of disaster. Consequently canoe-building was abandoned. It seems, then, that this technology carried within its behavioral by-product the seeds of its own destruction.

Strategy and its Paradoxes

Those of us who hope to see military technology ultimately wither and die derive some comfort from certain indications that it, too, carries within itself the seed of its own destruction. Some of these indications are obvious. If modern military technology is put to full use, it will destroy the industrial civilization which nurtures it and so will destroy itself. The comfort to be derived from this prospect is small. However, there is another, brighter prospect. The nemesis

of military technology may come from within itself, namely, from one of its principal nurturing sciences—the science of strategy.

In its most immediate application, that is, in actual combat, this science is called “tactics” and purports to embody the principles governing the development and use of war materiel in battle. The aims of strategy, however, reach much farther than the battlefield. Strategy purports also to embody the principles governing the use of battles (or their avoidance) in the prosecution of a war. Indeed, the scope of strategy in the last century or two has been still further broadened to include principles governing the use of war to prosecute national policy. It would seem, therefore, that strategy, if it lays claim to scientific status, should include in its scope not only such military subjects as tactics and logistics but also, in its more recent political context, economics, politics, geopolitics, sociology, etc.

Quite recently, war has become so uncomfortably dangerous, even for the global strategists, that it has become unrealistic to view war as Napoleon, Clausewitz, and Bismarck once viewed it, namely, as a convenient instrument of national policy. The talk nowadays revolves less around war itself and more around the *threat* of nuclear war as an instrument of national policy. Now *threat* is a psychological term, and having admitted such a term into its lexicon, the practitioners of strategy are obliged to throw an acknowledging nod to psychology.

It would be comforting for the strategists to enlist psychology in the service of war technology, and this can be done to a degree. The human psyche is tractable. Where some people exercise power over others, psychological know-how comes in handy for eliciting conforming behavior. But techniques of propaganda, indoctrination, and brainwash-

ing, while possibly useful in handling the living war material, are only peripheral to the central problem of strategy: the knowledge and, preferably, control of the psychic processes of those over whom one does not have power: the enemy.

Indeed, why should not psychology serve the strategists, if psychology purports to be a science of the psyche, and if everyone, including the enemy, has a psyche? I believe that the major block to an effective utilization of psychological knowledge in the service of strategy, as it applies to global conflict, is the ethos which underlies war technology and its supporting science. The ethos which underlies the practice of any science or technology, as we have seen, must justify the practice. Accordingly, the ethos which underlies the strategy of global conflict must justify the conflict. It is the ethos which assumes that, at least under certain circumstances, global conflict, including nuclear war, is justifiable. Spelled out, this is at times taken to mean (on our side) that nuclear war is preferable to the withdrawal of Western occupation troops from West Berlin or to the appearance of a Communist state in Latin America. The assignment of these preferences is usually justified on ideological grounds, often conveniently lumped into such a slogan as “Better dead than red.” Now it seems to me that an ethos which hides behind slogans will not stand up under psychological scrutiny, and for this reason, the practitioners of global strategy, if they wish to continue to practice their trade, must resist attempts to bring to bear psychological science on their problems.

Not only must the psychological validity of the better-dead-than-red axiom remain unquestioned in strategic circles, but also its relevance must remain unexamined. For the most part, global strategy, in order to be

taken seriously, must remain abstracted from real global issues. If issues are to be mentioned at all, they must be strategic issues, not substantive ones. The prizes of global strategy are simply strategic advantages, and the grand prize is something called "victory," which, according to Herman Kahn, can always be unambiguously defined. The victor, writes Kahn, is, by definition, the one who writes the peace treaty, much as the victor in a game of chess is the player who is entitled to say "checkmate."

It appears, therefore, that the paradigm of contemporary diplomatic-military policy is a vast game of strategy. In such a game, each decision can, in principle, be rationalized by the strategic advantage it confers. Each strategic advantage can, in turn, be rationalized in terms of its contribution to the probability of victory. But victory itself is not rationalizable. It is not legitimate to ask of what use is victory. To ask this question is to threaten the game itself and the *raison d'être* of the science of strategy. But such questions cannot be avoided if psychology is seriously to be brought into the picture. I believe this explains the striking psychological paucity of modern strategic theory.

In the context of the zero-sum game, to be sure, there is no need for psychology. *Homo economicus*, the mannikin invented by classical economics, the creature possessed by a single faculty, namely, greed, suffices as the model of the "rational player." His single tropism can well be built into a computer. His "rationality" is calculating ability. Since the theory of the two-person, zero-sum game is complete, a computer can, in principle, be preprogrammed to play any such game.

The Breakdown of Strategic Rationality in Non-Zero-Sum Games

The situation changes radically, however,

when we enter the realm of the non-zero-sum games. Thomas C. Schelling was one of the first students of strategy to point out the peculiar psychological problems which lurk in non-zero-sum games and which are neglected in formal game theory. A very elementary example will illustrate the type of problem. Suppose a dollar is to be divided between two people as follows. Each writes down independently of the other how much of the dollar he wants. If the sum of the two amounts does not exceed a dollar, each gets what he has asked for; otherwise the dollar is forfeited.

Schelling argues that each player should claim 50¢. His argument is intuitively justifiable and, incidentally, experimental evidence corroborates the expectation. This is one example of what Schelling calls the "prominent solution." An important goal in this game is to avoid the region in the game matrix where the sums exceed a dollar. Each player, being aware of this *common* interest, can contribute to the probability of avoiding the bad region by keeping his claim in the dollar low. He hopes the other is doing the same. On this assumption it is safe to claim 50¢, but not more, because each perceives that the other feels "safe" in claiming 50¢ and thus cannot be counted on to claim less.

Let us examine the matrix (Figure 1) of this game. For simplicity, the amounts to be claimed have been confined to 25¢, 50¢, and 75¢. The outcome (50, 50) is seen to be an equilibrium point, in the sense that neither player can independently depart from it without worsening his pay-off.

Let us now modify this game by changing just two entries, as shown in Figure 2. The equilibrium (50, 50) has now been destroyed, because now either player can improve his pay-off by moving away from the erstwhile equilibrium in one of the

		Player 2 claims		
		25¢	50¢	75¢
Player 1 claims	25¢	25, 25	25, 50	25, 75
	50¢	50, 25	50, 50	0, 0
	75¢	75, 25	0, 0	0, 0

FIG. 1.

		Player 2 claims		
		25¢	50¢	75¢
Player 1 claims	25¢	25, 25	25, 50	25, 75
	50¢	50, 25	50, 50	-25, 75
	75¢	75, 50	75, -25	0, 0

FIG. 2.

Columns and rows represent the respective claims of the players. The first of each pair of entries is the pay-off to the row player; the second, to the column player.

two directions. The most interesting feature of this game is that if both players pursue this course, both lose. An interesting psychological question is whether, in a situation of this sort, people will be deterred from seeking the additional advantage which moving away from equilibrium seems to confer.

The game shown in Figure 2 has another interesting feature. If one of the players has reason to suppose that the other will claim 75¢, he has a choice of three policies: he can stay with 50¢, he can follow suit with a counterclaim of 75¢, or he can "retreat" to 25¢. Obviously, in view of the resulting pay-offs, retreat is best, counterclaim is next best, and standing pat is worst. However, we have considered only the money pay-offs. We do not know what other pay-offs may be psychologically relevant. Retreat saves some of the money but, in some people's estimation, is hard on self-esteem, and, in repeated plays, may encourage further "aggression." The counterclaim is harder on the pocketbook but affords the satisfaction (for some people) of "punishing the aggressor." Staying with 50¢ and thus taking a loss of 25¢ is the choice most difficult to defend, because this choice seems to encourage the opponent to make the large claim and at the same time involves a loss for self. But its value as a *message* is not to be underestimated. In sticking to 50¢, the player is letting the other know that he is

firmly adhering to the principle of equity, neither retreating nor taking revenge but inviting the other to return to the social optimum (50, 50). It is a bid to call the *conscience* of the other into play.

All these are psychological, not game-theoretical, considerations. To be sure, it is an assumption of game theory that in principle all outcomes, together with their psychological overtones, are translatable into utility units. The idea is that, after such a translation has been made, the decision problem has been reduced to a game-theoretical one. But it is one thing to assume that something can be done "in principle" and quite another to do it. Moreover, even if utilities are definitely assigned, many non-zero-sum games do not have any intuitively acceptable solutions. The best known and simplest example of such a game is Prisoner's Dilemma, which we obtain if we delete the first row and the first column from the game shown in Figure 2, thus cutting off the "retreat" allowed in that game. An example of Prisoner's Dilemma is shown in Figure 3. We relabel the strategies C and D (for cooperation and defection). Strategy D dominates C for both players (i.e., C is best against either of the other's strategies). Yet the outcome DD is worse for both than the outcome CC.

I submit that these features of non-zero-sum games are quite suggestive of the sort

	C	D
C	50, 50	-25, 75
D	75, -25	0, 0

FIG. 3. Prisoner's Dilemma.

of dilemma which plagues international diplomacy and even military strategy of our day. I have also maintained that the rationales for decisions in those fields either make little or no use of psychology or, worse, use bad psychology—bad in the sense that the psychology which is used is more appropriate to “rational players” of classical (zero-sum) game theory than to human beings. However, the “rational player” psychology is, for the most part, inadequate in international affairs because the zero-sum-game models, often tacitly assumed to represent the issues, are grossly inadequate.

On the other hand, psychology cannot be avoided even in the oversimplified framework of present day diplo-military strategy, which depends so vitally on the credibility and effectiveness of threats.

An Experimental Approach

If psychological knowledge is to be used effectively in the guidance of policies and decisions, such knowledge must first be obtained. It cannot be obtained easily from actual experience in the world of diplo-military strategy because of the involvements of the participants, because of the complexities of the situation, because of the lack of opportunities to experiment or to conduct long-range observations, etc. We are thus reduced to the necessity of simulating the situations in the laboratory.

Now there are two ways of simulating a situation: one is by reproducing the features of its content; the other by trying to capture its psychological essentials. The first seems to have the advantage of realism. One can,

perhaps, view such simulation as a replica of the real state of affairs reduced in scale. The second has the advantage of simplicity and flexibility. The second method has one decided shortcoming. No extrapolation of the results obtained from it to any real situation can be seriously defended. But this is the case of every experiment *in vitro*. Extrapolations of the results to the situation *in vivo* are never safe. The limitation applies also to all “realistic” simulations, whether of a Strategic Air Command operations center or of the international game of nuclear diplomacy.

It goes without saying that proposals to learn about pertinent psychological processes from simulations, mock-ups, and games do not constitute proposals to rely on the findings in making decisions. The most one can expect from simulations is *hints* on what sort of psychological processes *may* be pertinent. The hope is that, in the light of these hints, the psychological problem lurking in diplo-military strategy might be reexamined. If only this could be accomplished, the laboratory procedures would already be strongly justified. The enemies of military technology hope for even more. They hope that the necessity of obtaining psychological knowledge will force the decision-makers to study psychology seriously, and that knowledge so obtained will reveal to them the folly of relying on military technology. In this way, the military complex may yet destroy itself before it destroys the rest of our civilization.

Of the two methods of laboratory approach, namely, realistic simulation and abstract formulation, we have chosen the latter. Some of the reasons for the choice have already been mentioned, namely, simplicity and flexibility of the procedures and theoretical tractability of the results. One other reason for our preference must be men-

9, 9	-10, 10	1, 1	-10, 10	1, 1	-10, 10	1, 1	-2, 2
10, -10	-1, -1	10, -10	-9, -9	10, -10	-1, -1	2, -2	-1, -1
Game 1		Game 2		Game 3		Game 4	

1, 1	-50, 50	5, 5	-10, 10	1, 1	-10, 10
50, -50	-1, -1	10, -10	-1, -1	10, -10	-5, -5
Game 5		Game 11		Game 12	

FIG. 4. Pay-off matrices.

tioned. It was our hope that abstractly formulated decision situations, deliberately separated from diplo-military content, would tap more basic psychological processes than realistic simulations. In a realistic simulation, a "player" might, for example, happily initiate a nuclear strike, because he knows it is only a game. Or, on the contrary, he might fail to initiate it (where a real decision-maker would), because the pressures operating on the latter cannot be reproduced in the laboratory. In an abstract game, only the structure of the game operates on the personality and on the motivation of the players. The trimming away of all the nonspecific components of conflict of interest contributes, one hopes, to bringing out the psychological essentials of such conflicts.

Our experiments, accordingly, were with games of the sort already mentioned, specifically, with the simplest of these, namely, the Prisoner's Dilemma type. We used, for the most part, seven of these games. The pay-off matrices are shown in Figure 4.

Our subjects were University of Michigan students. Each pair of subjects was used in only one experimental session, which consisted, in various experiments, of 300-700 consecutive plays of one or several games with communication between the players

disallowed. The pay-offs were in money, one mill per point.

A Summary of Experimental Results

So far we have conducted about 200 such experimental sessions, and our data consist of nearly 100,000 responses. Because of the great number of responses, rather stable statistical results have been obtained, which enable us to answer a number of questions about the behavior of university students in this situation. There is not time to describe in detail the experimental conditions, the results, or the mathematical models designed to describe the fundamentally stochastic process which underlies this form of interaction. I will conclude by listing the questions that can now be answered on the basis of the data so far gathered and analyzed. I will leave it to others to judge the psychological meaning of the answers, their import, and their possible relevance to the problems of arms control and disarmament. Only with regard to the last question will the connection between the answer and such problems be explicitly pointed out.

Q₁. What is the typical first response in a Prisoner's Dilemma game? (Note that purely strategic considerations dictate the response DD. Collective interest considerations dictate the response CC, but no

possibility for coordinating the latter response exists.)

A₁. The first response is practically random, as evidenced by the nearly equal proportions of CC, CD, DC, and DD responses. These proportions do not vary significantly in the seven games we have used.

Q₂. What happens thereafter in the short run?

A₂. The immediate trend which sets in is a decrease of C (the percentage of cooperative responses) shown by the decrease of a running average of C over ten consecutive plays. The extent of this decrease is such that the average C response over the first fifty plays is reduced from about 50 per cent to about 40 per cent.

Q₃. What happens in the long run?

A₃. After about fifty plays, the frequency of C responses averaged over the entire population of players steadily rises. This effect was not observed by earlier investigators who confined themselves to about thirty to fifty plays of Prisoner's Dilemma. The asymptotic value of C for the population is about 65 per cent.

Q₄. Is this asymptotic value also modal, i.e., characteristic of most of the player pairs in the population?

A₄. No. The modal frequencies are at the extremes. Most pairs either lock in on CC, i.e., achieve a tacit collusion to cooperate or, on the contrary, lock in on DD, giving up the possibility of achieving the biggest joint pay-off. The 65 per cent is only the average of a strongly bimodal distribution.

Q₅. Are the responses of the two members of a pair typically interdependent?

A₅. Definitely yes. Note that there are three possible answers to this question: (1) The responses of one player tend to elicit like responses on the part of the other. (2) The responses are independent. (3)

The responses of one player tend to elicit unlike responses in the other. All three possible answers can be rationalized in plausible terms. The first can be rationalized on the grounds of mutually stimulating attitudes—cooperation breeds cooperation and vice versa; the second, on the grounds of absence of explicit communication; the third, on the grounds that the biggest pay-offs come from responding noncooperatively to the other's cooperation.

If we assign the value 1 to each individual C response and 0 to each individual D response, the correlation coefficient of a run of paired responses will be positive if like elicits like, negative in the opposite case, and zero if the responses are independent. Empirically, we find this correlation coefficient distributed around the values of +0.5 to +0.6, indicating that the interdependence is of the first kind, i.e., that like responses are elicited more frequently than unlike ones.

Q₆. Is there other evidence of interdependency?

A₆. If we plot the probability of a CC response against the number of consecutive CC responses which have preceded it, we find that following a non-CC response the probability of a CC response (i.e., its "spontaneous" probability) is about 0.07 (averaged over the whole population of responses). On the other hand, following a *single* (first) CC response, the probability that the next one is also CC is about 0.60. Thereafter this probability continues to rise steadily with the number of consecutive CC responses which have already occurred, until the asymptotic value of about 0.95 is achieved. The situation is quite similar with the DD responses. A spontaneous DD response (i.e., following

a non-DD one) is comparatively rare with frequency of occurrence of about 0.10. Following a single DD response, the probability that the next one is also DD is 0.67. The probability keeps increasing with increasing number of previous consecutive DD responses until an asymptotic value of about 0.9 is reached.

These contingent probabilities constitute the most direct evidence of the "locking in" effect. Typically, a pair gravitates toward either CC or DD, and the probability of persisting in either state becomes the greater the longer the state has already persisted.

Note that this effect cannot be explained as a case of simple operant conditioning involving a selection of rewarded responses and elimination of punished ones. There are *pressures against* the persistence of either of the two states, CC and DD. The latter is directly punishing, while the former is associated with a temptation to defect in the hope of getting the biggest pay-off. The pressure against persisting in the unilateral cooperative states (CD or DC) is even greater, because these states are the most punishing for the unilateral cooperator. Nevertheless, there is evidence that these states also tend to persist, at least for a while.

Q₇. Is the behavior of player pairs different in games with different pay-offs?

A₇. Yes. Each of our games can be described by three parameters, representing (1) reward for cooperation, (2) temptation to defect, and (3) punishment for defection. We find that, if temptation and punishment are kept constant, C increases with reward; if reward and punishment are kept constant, C decreases with increasing temptation; if reward and temptation are kept constant, C increases with punishment, as one would expect.

Comparing Games 1 and 2 and Games 11 and 12 (cf. Figure 4) we see that rewards for cooperation are more effective in eliciting cooperation than numerically equal punishments for noncooperation (temptation being constant).

The differentiation of behavior in the different games is sharpest when each pair plays only one game 300 times in succession (as against other conditions when several games are scheduled in the same experimental run). Apparently when the same pair plays different games, the games "contaminate" each other. The games which ordinarily show a high degree of cooperation show a lower degree and vice versa.

Q₈. Aside from the persistence of CC and DD responses, are the other unilateral responses, CD and DC, equally persistent?

A₈. These tend to die out in the course of an experimental session. But long runs of unilateral responses (in exceptional cases, thirty to fifty) are at times observed. They can be naturally attributed to the attempts of one player to get the other to cooperate. We call such runs "martyr runs," because they are punishing to the well-intentioned player.

Q₉. How do martyr runs typically end?

A₉. We have compared the relative frequencies of CC and DD responses following a martyr run. We arbitrarily count three successive CC responses following a martyr run of at least three plays as a "conversion," while three consecutive DD's are counted as a "failure," i.e., the martyr gives up the attempt. Preliminary indications are that failures outnumber conversions about two and one-half to one.¹ One must remember, however, that an experimental session often includes

¹ Reversals also occur, that is, a conversion sometimes coincides with the martyr's giving up.

several such attempts, of which the last frequently succeeds. The fact remains that, in about one-half of all cases, an experimental session ends in a long run of cooperation.

Q₁₀. How are the lengths of the various runs distributed?

A₁₀. There is at least one noteworthy feature about these distributions. If we count all the runs of responses where CC does not occur (so-called conflict runs), we find that the proportion of such runs which have not yet been terminated at time t (measured by the number of responses) is equal to $e^{-a\sqrt{t}}$, where a is a parameter of the distribution.

Now very similar distributions are observed in the durations of strikes and of wars. One can interpret this result in two ways. One can assume that a general "law" is operating in the dynamics of conflicts which can be stated in qualitative terms thus: the longer a conflict lasts the harder it is to settle it. This is in view of the fact that the exponent of the decay rate is "weaker" than a linear function of time, which would be the case if the probability of termination of a conflict were independent of its duration. There is also another interpretation. Since our conflict runs were taken from a whole population of pairs, it is reasonable to suppose that the propensity for ending a conflict run is distributed in some fashion in the population of pairs. It turns out that, if this distribution is approximately logarithmic normal (like many distributions of parameters which take only positive values), then the observed rate of decay of the conflict runs would be just what we find it.²

² Strictly speaking, the distribution of propensities ought to be the inverse Laplace transform of $e^{-a\sqrt{t}}$, which is of the form $\frac{ae^{-a^2/4k}}{2\sqrt{\pi k^3}}$

Conceivably we can make a tentative decision in favor of one or the other hypothesis if we examine the distribution of runs of *single* pairs of subjects, thus avoiding the variation of the assumed "propensity to terminate conflict runs." If indeed the observed distribution of lengths of runs owes its form to the propensity distribution, then the distribution of lengths of runs in the responses of single pairs ought to be simply decaying exponentials (with different decay constants for different pairs) rather than the square-root exponentials which we observe in the whole population. Since the number of runs of a single pair is much smaller than in the whole population, the decision of whether the distribution is fitted better by an ordinary exponential or by a square-root exponential can be made only on the basis of a refined statistical analysis. This has not yet been done. The question of whether our data reveal something about the dynamics of conflicts or only something about the distribution in the population of a propensity to end a "conflict run" therefore remains open.

Now we come to the finding which perhaps has more direct bearing on arms control and disarmament. We ask what the probability is of ending a CC run. Recall that a pressure to end such a run is always present in the form of the temptation to defect. This temptation is tempered by the fear of punishment associated with DD, for it is natural to expect that defection will elicit retaliation and so the punishing state DD. If now we hold the numerical value of the punishment constant, how will the frequency of defections from the CC state be related to the size of the temptation?

Common sense would suggest that the relation would be a monotone one: the greater the temptation (other things being equal), the more probable the defection ought to be. Indeed, this argument is sometimes used to

suggest that a high level of armaments, properly controlled, makes for a more stable international situation than the state of disarmament, since the temptation to resume the arms race is so much stronger in the disarmed state. When everyone is disarmed, only a slight accumulation of, say, nuclear weapons makes a decisive difference in the balance of power, while in a highly armed, balanced state the discrepancy must be considerably higher in order to upset the balance.

We examine the probability of defecting from the CC state, once it has been achieved, in Games 1, 3, and 5. As can be seen from the matrix (cf. Figure 4), the punishment is the same in all three games. But the temptation is lowest in Game 1 and highest in Game 5. According to the common-sense hypothesis, then, on which the arms control argument also rests, defections from the CC state ought to be most frequent in Game 5 and least frequent in Game 1. Our data show that the probability of defection from CC is indeed lowest in Game 1. But it is considerably lower in Game 5 (the large-temptation game) than in Game 3 (the medium-temptation game). The respective probabilities of defection are 0.03, 0.12, and 0.05. Again there are two possible explanations. One is the natural selection principle. The pairs which achieved cooperation in the game with the highest temptation, namely, 5, are likely to be the pairs with the highest propensity for cooperation. Therefore they defect less often than the average pairs. The other explanation is that the achieving of cooperation in Game 5 is so difficult that, once it has been achieved, there is a strong reluctance to break the tacit agreement. If this is true, it has a bearing on the psychology of a disarmed world. Again we are not yet in a position to decide between the hypotheses (if, indeed, either is relevant). But the

possibility that some such psychological forces may be operating is certainly not excluded.

Concluding Remarks

It behooves us, I think, to pay attention to such psychological considerations when we analyze the world situation with a view of weighing the various proposals aimed at minimizing the danger of war. It is unfortunate that so much ingenuity is directed at the invention of elaborate mechanical safeguards and at purely strategic analyses, and so little effort is undertaken with a view toward understanding what makes people behave as they do. The strategists tend to forget that decision-makers are human beings, often, alas, an all-too-understandable mistake, but nevertheless a dangerous one. If a trivial game for pennies turns out to be such a rich source of information, revealing intricate and stable patterns of human behavior in mixed-motive situations, how much more there is to learn in real life. Let us never forget that the most important question remains unanswered—a question of supreme importance to us as human beings. How did we get into this bind, where the most unspeakable crimes against humanity are not only thinkable but are actually the subject of an on-going serious discussion? What we ought to do is find out how to make the unspeakable unthinkable and to keep it so. This task involves not something that one does with hardware nor something that one calculates on a computer. It involves the acquisition of substantive knowledge of the human psyche.

The laboratory experiments yielding “hard data,” such as those we have described, should not be considered as sources of real knowledge about the human psyche but only as pointers to a tortuous road toward such knowledge.