COMMENTS AND REPLIES

COMMENT ON KOLLOCK, *ASR*, DECEMBER 1993

COOPERATION UNDER UNCERTAINTY: WHAT IS NEW, WHAT IS TRUE, AND WHAT IS IMPORTANT*

Jonathan Bendor
Stanford University

Roderick Kramer
Stanford University

Piotr Swistak
University of Maryland

William James once lamented that critics of his theory of pragmatism used the "stock phrase that 'what is new is not true, and what is true is not new' " (quoted in Merton 1967:21–22). James may well have been in the right in that particular controversy, but the "stock phrase" he quotes does capture an important fact: To count as a genuine contribution to science, a finding should be both true *and* new—and, better yet, significant as well. Unfortunately, in the case of Kollock's 1993 study (henceforward Kollock) of the iterated Prisoner's Dilemma (IPD) under monitoring uncertainty ("noise"), the stock phrase turns out to be all too accurate. Much of what is true (and significant) in Kollock's paper is not new, and much of what is new is either not true or not significant.[1] These are strong claims, and they must be substantiated. Thus, we begin with what is true and significant, but not new: the paper's problem, method, and its main conclusion.

* Direct correspondence to Jonathan Bendor, Graduate School of Business, Stanford University, Stanford, CA 94305-5015 (fbendor@gsb-lira.stanford.edu). We thank Jim Baron, Mayer Zald, and, especially, Joel Podolny for their very helpful comments.

[1] There is one exception, however. Toward the end of this note we point out one finding—the success of "stingy" strategies in situations of intense uncertainty—that seems to be significant, quite possibly true, and an original discovery.

TRUE AND SIGNIFICANT, BUT NOT NEW: THE PROBLEM, THE METHOD, AND THE MAIN CONCLUSION

First, Kollock does not address a new problem. What he identifies as "the key issue . . . , [namely] the compromise that must be made between vulnerability to exploitation . . . and vulnerability to needless cycles of recrimination . . ." (p. 769), was clearly identified as a problem by Axelrod a decade ago (1984: 183). Describing the essence of the problem, Axelrod and Dion (1988) wrote in a review article ". . . for larger amounts of noise, there is a trade-off: unnecessary conflict can be avoided by generosity, but generosity invites exploitation" (p. 1387).

Second, if the problem is not new, then perhaps the method of solving it is? Yet Kollock's method—a computer simulation of a round robin tournament of strategies playing the IPD—is not new either. Since the idea of simulated tournaments was pioneered by Axelrod (1980a, 1980b), such tournaments have been run by many others; indeed, several have been run to address the very same problem (e.g., Axelrod 1984; Donninger 1986; Bendor, Kramer, and Stout 1991; Nowak and Sigmund 1992).

And finally, if neither the paper's problem nor its method are new, then perhaps its central conclusion is? Again, the answer is "no": Kollock's main result is not new either. His main finding is that when playing the noisy IPD, "strategies that use a more relaxed accounting system than TIT FOR TAT often have important advantages" (p. 784). This is probably true, and we agree with him that this is significant; but it is not new, having already been reported in an earlier study (Bendor et al. 1991). The main finding of Bendor et al.'s noisy IPD tournament was that : ". . . TFT performed rather poorly. In contrast, strategies that were generous (i.e., cooperated more than their partners did) were quite effective" (p. 691).[2] The concepts

[2] A similar effect was also observed in a simulated tournament by Donninger (1986). The evo-

of generosity and of relaxed accounting systems turn out to be very similar: All of the strategies that Kollock identifies as having relaxed accounting systems are generous, as Bendor et al. define the term; those he identifies as having the most restrictive accounting systems are stingy in Bendor et al.'s framework. Further, the nature of this advantage—that generosity (relaxed accounting systems) works by dampening unintended vendettas—was identified by Bendor et al. as well (1991:706).[3]

WHAT IS NEW BUT PROBLEMATIC

Lack of originality is a significant weakness, but a study may still be valuable as a source of new perspectives or observations, or even as a replication of earlier results. Are, then, Kollock's new observations meaningful? And how much validity does his study add to what already had been observed? We think the answer to the first question is "no" (except for the point of footnote 1), while the answer to the second is "some validity, but less than might have been the case." And this is because, in addition to its lack of novelty, Kollock's study suffers from two major errors. One is substantive; the other is methodological.

Substantive Problems: Understanding the Effects of Noise

To see why Kollock's observations regarding the effects of noise cannot be taken quite at face value, they ought to be reconsidered in the context of what is known. Kollock portrays uncertainty as having only negative consequences. While these undoubtedly exist—as has been known for some time (Downs, Rocke, and Siverson 1985; Molander 1985; Bendor 1987; Mueller 1987)—uncertainty's *benign* effects may well be more important (Boyd 1989; Bendor 1993). The benign effects of noise concern the two most fundamental aspects of the evolution of cooperation: its *emergence*—how cooperation can get started in a world of suspicious egoists—and its *stability*—once established, how cooperative strategies can resist "invasion" by less cooperative ones. We consider these in order.

How cooperation begins is the first fundamental issue in the evolution of cooperation. Hence, we must understand how noise can help or hinder these beginnings. Though technically correct, Kollock's observation on how noise can facilitate the emergence of cooperation in ecologies of certain "not-nice" strategies, such as SUSPICIOUS–TIT FOR TAT (a strategy in his simulation), is so understated that it borders on inaccurate. While he notes that for strategies like SUSPICIOUS-TFT "a little noise is a boon because an occasional distortion in a positive direction can lead to [cooperation]" (p. 776), something much stronger can be shown: This benign effect is not restricted to situations with just "a little noise." Indeed, it has been proven (deductively) that for a very wide class of shocks, *any* amount of noise leads an ecology of SUSPICIOUS-TFT to exhibit as much cooperation in the long run as would a TFT ecology (Bendor 1993:731). In either case, half of all moves would be cooperative. This is a big improvement over the Hobbesian condition that would prevail in a noiseless ecology of SUSPICIOUS-TFT, where none of the moves would be cooperative. Thus if decision-makers believe it is too risky to initiate cooperation, but are willing to reciprocate it (a plausible circumstance), noise can be very helpful. Thus, the convergence between SUSPICIOUS-TFT and TFT reported in Kollock's Figure 1 is no accident—he is simply observing a specific instance of a general result (Bendor 1993).

The second fundamental issue in the evolution of cooperation is what happens when a cooperative strategy becomes successful and takes over an entire population. In this stage it is crucial to show that such a population is in equilibrium (i.e., that a common cooperative strategy is evolutionarily stable).

---

lutionary advantage of generous strategies was also shown in Nowak and Sigmund's (1992) simulation.

[3] Indeed, the essence of this solution goes back further: "What is needed to escape from a destructive vendetta is some unconditional generosity" (Molander 1985:612–13; also see Mueller 1987: 714). It is true, as Kollock states, that in the studies by Molander and Mueller the strategies were constrained in every period to a binary choice of cooperate or defect, whereas Kollock allowed for degrees of cooperation. But strategies that can exhibit degrees of cooperation are not new either; those in Bendor et al. had this property.

In this matter Kollock commits two errors: First, he attempts to study stability inductively (something that cannot be done); second, he offers insights and conjectures that are known to be false.

To understand the effects of noise on evolutionary stability, we must first understand which strategies are stable (and in what sense they are stable) in noiseless ecologies. Only then can we make comparative judgements about whether noise is beneficial or harmful. Whereas Kollock portrays noiseless environments as being conducive to cooperation, it turns out that the evolutionary stability of strategies in the noiseless IPD is riddled with problems, far more so than a casual reading of Axelrod (1984) might suggest. As we will see, whereas these problems exist in games without noise, *they vanish when noise is introduced.*

The first problem of stability in noiseless games relates to Kollock's claim that "In the absence of noise, . . . once TIT FOR TAT has established itself and created a world of mutual cooperation, it cannot be invaded by other strategies" (p. 782). *This claim is false.* As Selten and Hammerstein (1984) pointed out a decade ago, any population of TFT can be invaded by a single ALWAYS-COOPERATE (ALL-C). Indeed, *any* single nice strategy (one that is never the first to defect) can invade an ecology of TFT. The reason is simple: Without noise, all nice strategies are observationally equivalent, if only nice strategies are in the ecology. Thus if ALL-C invades a population of TFT, everyone cooperates with everyone else in every period, so both strategies do equally well.[4] If the concept of stability connotes a return to the earlier status quo, no evolutionarily stable strategies exist in the noiseless IPD. Interestingly, however, noise eliminates this problem: Evolutionarily stable strategies *do* exist when noise is present. Again the reason is simple: With noise present, ALL-C and TFT are no longer behaviorally indistinguishable.[5] Noise provides an essential benefit because it enables strategies to discriminate among each other, which ensures the existence of evolutionary stable strategies (Selten 1983).

The second problem concerning stability in the noiseless IPD (discovered by Boyd and Lorberbaum 1987) is that not only can invading mutant strategies do as well as TFT, as is the case with ALL-C, but they can even outscore TFT. And this can happen regardless of how few mutants invade.[6] This observation by Boyd and Lorberbaum implies that under some evolutionary processes, TFT can be eliminated by the more successful mutant. Interestingly, this form of instability also disappears when noise is introduced (Boyd 1989).

All of the above results are deductive. This means that *we do not need simulations to look for strategies that are evolutionarily stable in noisy games*. Stability can be established analytically, and stable strategies have already been identified (Boyd 1989). We should note that among evolutionarily stable strategies there are some with "nice" properties (e.g., Boyd 1989) in that they cooperate infinitely often with their clones, and as the probability of misperception goes down, the rate of cooperation goes up. Thus, these strategies can uphold a structure of cooperation against nice strategies that are excessively soft (e.g., ALL-C) as well as against nasty strategies that do not cooperate (e.g., ALL-D).

Methodological Issues

If evolutionary stability can be studied deductively, then using simulations to study it is inappropriate. Accordingly, a section of Kollock's paper titled "The Evolutionary Stability of Strategies" presents a misleading argument. In this section Kollock first dis-

[4] This means that TFT cannot be stable in a strong sense (i.e., it cannot ensure that all invading mutant strategies will die out or equivalently that the equilibrium from the pre-invasion stage will be restored). Other, weaker forms of stability are possible, however (Bendor and Swistak 1995).

[5] There is an interesting parallel here with the study of deviance and social control. Theorists have long argued that deviance serves a useful social function, partly because it allows sanctions to be demonstrated. Without noise, ecologies of nice strategies never exhibit deviations, so sanctions are never manifested.

[6] In Boyd and Lorberbaum's (1987) example, a population of TFT's is invaded by arbitrarily few TIT FOR TWO TATS (defect in response to two consecutive defections of your opponent) and SUSPICIOUS-TFT. It is simple to check that TF2T will outscore TFT if "the shadow of the future" (Axelrod 1984) is sufficiently long.

cusses the notion of an evolutionarily stable strategy; he then analyzes which of the seven strategies used in the simulation can be invaded by the other six. Given the title of the section and the structure of the presentation, the reader is clearly led to believe—though Kollock never states it explicitly—that his simulation findings pertain to the evolutionary stability of some of these seven strategies. Hence it is implied that the finding that "TAT+1 is . . . very resistant to being invaded" (p. 782) says something about its stability.

Do, then, Kollock's results (Table 1) have anything to do with the evolutionary stability of his strategies? In fact, they do not. Indeed, it is easy to prove that neither TFT, SUSPICIOUS-TFT, TAT+1, TAT−1, nor CYCLE is stable. Moreover, the most successful strategy in Kollock's simulation, TAT+1, is unstable at *any* level of noise.[7] Inductive inferences, like those based on his simulation, can be misleading; deduction is a good critical companion to such analyses. In general, simulation results cannot prove that a strategy is evolutionarily stable—with or without noise. The essence of a stable strategy is that it cannot be invaded by *any* mutant strategy, and this *can only be established analytically*.

Nevertheless, while simulation is an inappropriate method for identifying stable strategies, it can be used to explore the out-of-equilibrium dynamics that arise when strategies are not equally fit. In these situations, it is not presumed that any of the strategies in the tournament are evolutionarily stable. Instead, the issues are more modest: Which kinds of strategies perform comparatively well in the ecology at hand, and what can be inferred from these observations? Here Kollock makes two contributions. (For reasons we will make clear, we think the first contribution is more significant than the second; indeed, problems associated with the second may detract from the first.)

First, unlike most previous noisy tournaments (e.g., Donninger 1986; Bendor et al. 1991), Kollock's simulation varies the amount of uncertainty (noise), which yielded a very interesting finding: When all seven strategies were included in the simulation, the two *stingiest* strategies placed first and second in the tournament at the highest levels of noise (Figure 6, p. 781). Indeed, though Kollock considers the finding supporting the advantages of generosity to be his central result (see his abstract, p. 768), we believe his most important original contribution is, in fact, that there are *dis*advantages of generosity in situations of intense uncertainty. This finding deserves further research because of its significance. It also requires further research because the effect appears to be ecologically sensitive: Stingy strategies did not dominate at high noise levels when there were either five or six strategies in the tournament (Figures 4 and 5, pp. 779 and 780). And this raises the issue of the mix of strategies Kollock uses and how he chose them.

Earlier studies showing the generally benign effects of generosity were based on a limited set of observations (i.e., ecologies). Potentially, then, Kollock's second contribution is his investigation of a new ecology. Unfortunately, the method he used to select strategies departs from standard practice, and his paper suffers thereby: It is unclear what, if any, generalizations can be drawn from Kollock's simulations. A set of objects used for a simulation resembles a sample from which statistical inferences are made—a sample is useful only if it is representative (i.e., random, unbiased) of the population from which it is drawn. If we are to make reasonable inferences, objects chosen for simulations (in this case, strategies) should be selected systematically.

Unlike Axelrod (1980a, 1980b), Kollock did not solicit strategies from different strategists; all of the strategies in his study are of his own choosing. This procedure runs the risk of idiosyncratic choice. Naive selection may produce a homogeneous sample, skewed by well-known cognitive biases.[8] Such a sample of strategies would not be representative of any population of strategies—neither theoretical populations (i.e., of program-

[7] A formal proof of these claims was included in an earlier version of this comment. The proof was made available to the *ASR* reviewers and to Peter Kollock.

[8] These include the confirmation bias, egocentric bias, and experimenter expectancy. See Gilovich (1991) for a general treatment of how these biases influence decision makers' abilities to generate representative samples and tests of hypotheses.

mable strategies, admissible in computer tournaments) nor empirical ones (strategies played by people in real ecologies). Unfortunately, the homogeneous and skewed nature of Kollock's sample is remarkably evident: Most of the strategies are minor variants of TFT. Thus, because Kollock makes claims about the population of generous (programmable) strategies, and because these claims are inductive inferences based on a small, homogeneous, and probably biased sample, his study's design weakens the internal validity of his claims about the benefits of less restrictive accounting systems in simulations of the IPD. For similar reasons, his method of generating a sample of strategies weakens the study's external validity as well. Hence, even as a replication of earlier studies Kollock's study is of limited value—it is a weaker replication than it might have been.

Compare Kollock's sampling method with Axelrod's (1980b). In his second tournament, Axelrod elicited strategies from 63 participants in six countries, thus obtaining a far more heterogeneous mix of strategies. A reasonable claim can be made that these strategies came from a representative sample of a population of sophisticated players. Axelrod's argument about TFT's robustness would have been much less persuasive had he run a simulation in which TFT competed against a handful of mostly similar alternatives that he had arbitrarily constructed. Following Axelrod's lead, recent studies investigating TFT's robustness in noisy tournaments have exhibited more ecological variety than does Kollock's simulation (Donninger 1986; Bendor, Kramer, and Stout 1991; Nowak and Sigmund 1992).

Hence, where one would expect simulation to exploit the computational powers of computers to study a rich ecology of diverse strategies, Kollock's simulation does not do it. It is thus not a reliable replication of an earlier finding that generosity can promote cooperation under noise. Conversely, it tries to use the computer in situations where simulations cannot work at all—in the study of evolutionary stability. Thus, methodologically Kollock's study is in the worst of all possible worlds: it applies the wrong tools to analyze some issues, and where it is appropriate to use these tools it does not exploit their full power.

Jonathan Bendor *is Professor of Political Economics at the Graduate School of Business at Stanford University. His research focuses primarily on the evolution of cooperation, models of boundedly rational behavior, and theories of organizational decision-making.*

Rod Kramer *is Associate Professor of Organizational Behavior at the Graduate School of Business at Stanford University. His research focuses primarily on decision-making in conflict situations, such as social dilemmas, negotiations, and international disputes. He is the co-editor of two recent books,* Negotiation as a Social Process *(with David Messick, Sage, 1995) and* Trust in Organizations *(with Tom Tyler, Sage, 1996).*

Piotr Swistak *is Associate Professor of Political Science at the University of Maryland in College Park. His current interests include theories of choice, foundations of social and political institutions, and the methodology and philosophy of social science.*

REFERENCES

Axelrod, Robert. 1980a. "Effective Choice in the Prisoner's Dilemma." *Journal of Conflict Resolution* 24:3–25.

———. 1980b. "More Effective Choice in the Prisoner's Dilemma." *Journal of Conflict Resolution* 24:379–403.

———. 1984. *The Evolution of Cooperation.* New York: Basic Books.

Axelrod, Robert and Douglas Dion. 1988. "The Further Evolution of Cooperation." *Science* 242:1385–90.

Bendor, Jonathan. 1987. "In Good Times and Bad: Reciprocity in an Uncertain World." *American Journal of Political Science* 31:531–58.

———. 1993. "Uncertainty and the Evolution of Cooperation." *Journal of Conflict Resolution* 37:709–34.

Bendor, Jonathan, Roderick Kramer, and Suzanne Stout. 1991. "Cooperation in a Noisy Prisoner's Dilemma." *Journal of Conflict Resolution* 35:691–719.

Bendor, Jonathan and Piotr Swistak. 1995. "Types of Evolutionary Stability and the Problem of Cooperation." *Proceedings of the National Academy of Sciences* 93:3596–3600.

Boyd, Robert. 1989. "Mistakes Allow Evolutionary Stability in the Repeated Prisoner's Dilemma Game." *Journal of Theoretical Biology* 136:47–56.

Boyd, Robert and Jeffrey Lorberbaum. 1987. "No Pure Strategy Is Evolutionarily Stable in the Repeated Prisoner's Dilemma Game." *Nature* 327:58–59.

Donninger, Christian. 1986. "Is It Always Efficient To Be Nice? A Computer Simulation of Axelrod's Computer Tournament." Pp. 123–34 in *Paradoxical Effects of Social Behavior*, edited by A. Diekmann and P. Mitter. Heidelberg, Germany: Physica-Werlag.

Downs, George, David Rocke, and Randolph Siverson. 1985. "Arms Races and Cooperation." *World Politics* 118–46.

Gilovich, Thomas. 1991. *How We Know What Isn't So*. New York: Free Press.

Kollock, Peter. 1993. "'An Eye for an Eye Leaves Everyone Blind': Cooperation and Accounting Systems." *American Sociological Review* 58:768–86.

Merton, Robert. 1967. *On Theoretical Sociology*. New York: Free Press.

Molander, Per. 1985. "The Optimal Level of Generosity in a Selfish, Uncertain Environment." *Journal of Conflict Resolution* 29:611–18.

Mueller, Ulrich. 1987. "Optimal Retaliation for Optimal Cooperation." *Journal of Conflict Resolution* 31:692–724.

Nowak, Martin and Karl Sigmund. 1992. "Tit for Tat in Heterogeneous Populations." *Nature* 355:250–53.

Selten, Reinhard. 1983. "Evolutionary Stability in Extensive 2-Person Games." *Mathematical Social Sciences* 5:269–363.

Selten, Reinhard and Peter Hammerstein. 1984. "Gaps in Harley's Argument on Evolutionary Stable Learning Rules and in the Logic of 'Tit For Tat.'" *The Behavioral and Brain Sciences* 7:115–16.

COMMENT ON KOLLOCK, *ASR*, DECEMBER 1993

COOPERATIVE STRATEGIES IN LOW-NOISE ENVIRONMENTS*

Edward B. Reeves
Morehead State University

Timothy C. Pitts
Morehead State University

Kollock (1993, henceforward Kollock) employed computer simulations permitting a diversity of cooperative strategies in repeated games. He then made an important claim: In "noisy" environments (i.e., environments in which an actor's contribution has some nontrivial probability of being misperceived by an alter) a generous, forgiving strategy is more successful than one that follows the Old Testament injunction of "an eye for an eye." In effect, Kollock's view takes exception to Axelrod's (1984) touting of TIT FOR TAT as the omnibus game strategy.

In Kollock's simulations, the generous strategy TAT+1 was extremely successful (see especially his Table 1, p. 782). TAT+1 was able to invade five of the other six strategies over nearly the full range of noise frequencies. At the same time it could not be invaded by any other strategy except TF2T-MAX, for which the outcome was random. TIT FOR TAT, on the other hand, could only invade less generous strategies, such as SUSPICIOUS-TFT, TAT–1, and TF2T-MIN (the last only when the noise frequency was less than 20 percent). Meanwhile, TIT FOR TAT was invaded by more generous strategies, such as TAT+1 and TF2T-MAX. Punitive strategies such as TF2T-MIN and TAT–1 were generally unable to invade more generous strategies (TAT+1, TF2T-MAX, TIT FOR TAT), while they themselves could be invaded by these same strategies.

The implications of Kollock's analysis are striking. When substantial noise is present, we should expect the evolution of stable strategies of generous cooperation that cannot be displaced by less generous strategies. In this comment, we question if Kollock's conclusion is consistent with the mixture of successful strategies that characterizes real-world social exchange, even in the real-world examples cited by Kollock as being most hospitable to generous, forgiving strategies.

Kollock offered several examples of social environments that favor a generous, forgiving strategy. For instance, he offered the case of food exchange in hunting and gathering societies. Our reading of the anthropological literature in this area found that in small-scale societies people are sometimes generous with their food resources, especially when kinship ties are close. Otherwise, however, more strict accounts are kept, and outright stinginess is not unknown. Thus, we found evidence of exchange strategies characterized in the literature as "balanced reciprocity" and "demand sharing" (a proactive TIT FOR TAT strategy), as well as evidence of repetitive bickering over failure to meet obligations

* Direct correspondence to Edward B. Reeves, Department of Sociology, Social Work and Criminology, Morehead State University, Morehead, KY 40351 (e.reeves@morehead-st.edu).