

Surprising Ramifications of the Surprise Exam Paradox*

Steven O. Kimbrough
565 JMHH
kimbrough@wharton.upenn.edu
215-898-5133

PAMLA talk, 17 March 2003

*File: surprise-exam-foils.pdf.

A word on methodology

Philosophy:

- Ordinary language: What is done by retired, over-the-hill intellectuals, propagandizing pundits, and middle-aged men drinking beer in dark bars.
 - Sociology: What is done by the people employed in the various Philosophy departments.
- ⇒ Ordinary language: A dogged attempt to think clearly on certain fundamental matters.

How to Do Philosophy

“A certain body of indefinable entities and indemonstrable propositions must form the starting-point for any mathematical reasoning; and it is this starting-point that concerns the philosopher. When the philosopher’s work has been perfectly accomplished, its results can be wholly embodied in premisses from which deduction may proceed. Now it follows from the very nature of such inquiries that results may be disproved, but can never be proved. This disproof will consist in pointing out contradictions and inconsistencies; but the absence of these can never amount to proof. All depends, in the end, upon immediate perception; and philosophical argument, strictly speaking, consists mainly of an endeavour to cause the reader to perceive what has been perceived by the author. The argument, in short, is not of the nature of proof, but of exhortation.”

—Bertrand Russell, *The Principles of Mathematics*, 1902, XV, 124

(See also, “The Ways of Paradox” by W.V. Quine.)

The Surprise Examination Paradox

- Aka: Surprise Hanging Problem
- “There will be a surprise exam given in one of the next 6 meetings of the class.”
- Reasoning by backwards induction. . .

From Grim et al., *The Philosophical Computer*, page 163

The similarity of this reasoning to that of the argument for dominant defection throughout a series of known finite length is worth noting because of course the Surprise Examination is treated standardly in the philosophical literature as a *paradox*, thought to hide some fallacious piece of logical legerdemain. That the same form of reasoning is thought of as valid in the theoretical economics literature, though perhaps inapplicable in some practical sense, indicates that important work remains to be done in bridging the two bodies of work.

First question on the exam

1. Explain the fallacy in the reasoning that led you to believe it impossible for me to give you a surprise exam as announced.

Elementary confusion by the students: Speech alone (in this context) does not have the power to prevent an exam from being given, even a surprise exam.

The real question: Can the teacher give a surprise exam *and* speak truly in saying that there will be a surprise exam?

Will try to reconstruct and present a proper way of reasoning about this.

Begin with the one-shot problem: “There will be a surprise exam tomorrow.”

Version 1 Applied to the One-Shot Surprise Exam Problem

1. $E \vee \neg E$

There will be an exam tomorrow or not, a tautology.

2. $l = \frac{2}{3}$

Arbitrary threshold (line or level); may be changed without loss of generality.

3. $P(E) \geq l \rightarrow \neg S$

If the probability of an exam tomorrow is greater than or equal to the stipulated threshold, then no surprise.

4. $V(a) \rightarrow (E \wedge S)$

There will be an exam tomorrow and it will be a surprise, asserted by the teacher. $a =$ the assertion by the teacher that there will be an exam on the next class (E) and that it will be a surprise (S). If the assertion is truthful or veridical, $V(a)$, then $(E \wedge S)$.

5. $E \rightarrow P(E) = 1$

If there is an exam tomorrow, then the probability of that there is an exam tomorrow is 1.

6. $P(E) = 1 \rightarrow P(E) \geq l$

A simple mathematical truth.

$$\vdash (\neg S \wedge \neg V(a)) \vee \neg V(a)$$

Comment: Valid, but unsound. Premise (5) is the problem.

Version 2 of the Students' Reasoning Applied to the One-Shot Surprise Exam Problem

1. $E \vee \neg E$

2. $l = \frac{2}{3}$

3. $P(E) \geq l \rightarrow$
 $\neg S$

4. $V(a) \rightarrow$
 $\Box(E \wedge S)$

Comment: Valid, but now premise
(4) is problematic.

5. $\Box E \rightarrow$
 $P(E) = 1$

6. $P(E) = 1 \rightarrow$
 $P(E) \geq l$

$\vdash \neg V(a)$

Version 3 of the Students' Reasoning Applied to the One-Shot Surprise Exam Problem

$$1. E \vee \neg E$$

$$2. l = \frac{2}{3}$$

$$3. P(E) \geq l \rightarrow \neg S$$

$$4. V(a) \rightarrow (P(E) = 1 \wedge S)$$

$$5. P(E) = 1 \rightarrow E$$

$$6. P(E) = 1 \rightarrow P(E) \geq l$$

$$\vdash \neg V(a)$$

Comment:

Now premise (4) is problematic. If this *is* what our teacher meant, then we'll simply get another teacher, who will mean something else. The real question is whether when we find such a teacher she can speak truly and give the surprise exam.

Version 4 of the Students' Reasoning Applied to the One-Shot Surprise Exam Problem

$$1. E \vee \neg E$$

$$2. l = \frac{2}{3}$$

$$3. P(E) \geq l \rightarrow \neg S$$

$$4. V(a) \rightarrow (E \wedge S)$$

$$5. P(E) = 1$$

$$6. P(E) = 1 \rightarrow P(E) \geq l$$

$$\vdash \neg V(a)$$

Comment:

Again, the argument is valid and the key premise is (5). Its justification is that there's no where else to put the probability mass. There will be an exam and there is only one day available for it, so all the probability has to be on that day. But this is wrong. $P(E)$ can in principle be anything at all.

Version 5 of the Students' Reasoning Applied to the One-Shot Surprise Exam Problem

1. $E \vee \neg E$

2. $l = \frac{2}{3}$

3. $(P(E) < l \wedge E) \rightarrow S$

4. $V(a) \leftrightarrow (E \wedge S)$

5. $P(E) < l$

$\vdash (E \wedge S \wedge V(a)) \vee (\neg E \wedge \neg V(a))$

Bingo!

Notice that assumption (4) has been strengthened to a biconditional. This is harmless and could have been done for the earlier versions. The strengthening amounts to accepting a rule that credits the teacher with speaking truthfully, $V(a)$, if what she said—that $(E \wedge S)$ —is in fact true.

The upshot: In the one-shot surprise exam problem, the teacher must either speak falsely (e.g., by making a self-contradictory statement) or speak truly but with a probability no larger than l . Only by putting herself at risk of falsehood is it possible for her to speak truly in this case. By taking a risk (of speaking falsely) the teacher expands her scope of action.

The teacher can trade risk (of falsehood) for reward (being able to give a surprise exam).

Now the 6-shot (n-shot) Surprise Exam Problem

If the teacher is willing to run a risk, $r > 1 - l$ of speaking falsely, then it is not true that surprise could not lurk on the last day. This suffices to block the backwards induction and to undue the students' reasoning in the n-shot case.

But. . . The soundness of version 5 relies on the teacher being willing to accept a risk of at least $1 - l$ of speaking falsely. Given this, many would choose not to utter the one-shot surprise exam assertion. Honesty, integrity, prudence, or whatever may well prevent a reasonable person from saying something they know to have a chance higher than $\frac{1}{3}$ of being false. Better to keep silent.

What if the students know this?

Students' Reasoning Applied to the Augmented One-Shot Surprise Exam Problem

1. $E \vee \neg E$

2. $l = \frac{2}{3}$

3. $(P(E) \geq l \wedge E) \rightarrow \neg S$

4. $V(a) \leftrightarrow (E \wedge S)$

5. $P(E) \geq (1 - r)$

6. $r < (1 - l)$.

$$7. (P(E) \geq (1 - r) \wedge r < (1 - l)) \rightarrow P(E) \geq l$$

$$\vdash (\neg V(a) \wedge E \wedge \neg S) \vee (\neg E \wedge \neg V(a))$$

Teacher's falsehood, validly deduced.

A deeper lesson lurks

More shots attenuate the teacher's verisimilitude scruples. Sufficient is:

1. In each period, initially the probability of the exam is less than l , and
2. The total probability of having the exam is greater than or equal to $(1 - r)$

This is trivially arranged for any l , and for any $(1 - r) < 1$, provided enough periods are available. Simply decide to give the exam with equal probability to every period. Again concretely with the l and $(1 - r)$ values given above, give a probability of $\frac{1}{5}$ of holding the exam on each of 5 days. The exam is held. If the exam is held on the fifth day, that morning the students will

know that the exam will be held that day (assuming they are certain the exam will be held at all). There is only a 1 in 5 chance this will happen, which is above $(1 - r)$ and acceptable to the teacher. If the exam occurs on day 4, the students have a 50% certainty that morning, which is below l . And earlier is even better for the teacher's veracity. Further, with enough periods the teacher can set the probability of giving the exam to 1 and still have a probability $< r$ of not surprising the students, as we have just seen in the example.

In sum on the Surprise Exam Paradox

The n -period case generalizes the 1-shot case. The teacher can speak truly in this form provided the teacher is willing to undertake some risk, $r > 0$, of speaking falsely and providing n is large enough (given r and l). The teacher cannot be certain of speaking truly, but in this respect the case is like most. Usually, when we assert we take some chance of speaking falsely, even with the best of intentions. What is odd is to interpret a speaker otherwise. The only way the teacher could have spoken truthfully and given the surprise exam was to have spoken with some chance of speaking falsely.

The students erred in presuming incorrectly regarding the teacher's toleration of risk.

The Point Again

A = “Here’s my randomized choice mechanism.”

B = “There will be an exam and it probably will be a surprise.”

C = “There will be an exam and it will be a surprise.”

1. A+say A + say B \Rightarrow no paradox.
2. A+say A + say C \Rightarrow no paradox; obvious that teacher simply risks error.
3. A+ say C \Rightarrow no paradox; less obvious—but still true—that teacher simply risks error.

And Again

Let $G(i)$ be the situation with i games left to play. To make an induction argument, we need to show two things:

1. Base step: If $G(1)$, then Condition (no surprise exam, i.e., $\neg(E \wedge S)$).
2. Induction step: If Condition (no surprise exam) at $G(n)$, then Condition (no surprise exam) at $G(n + 1)$.

I have argued that *both* the base step and the induction step fail (under plausible conditions).

Question 2 on the Exam

2. In a 100-shot Repeated Prisoner's Dilemma game, played between the teacher and an unknown, but fully competent human subject, the teacher announces that she will gain the reward from mutual coöperation at least 2 times, net. That is, if P is the penalty for mutual defection and R is the reward for mutual coöperation, the teacher is asserting that she will get at least $98 \cdot P + 2 \cdot R$ points from the 100 trials. Can this assertion be plausibly justified? Why or why not?

The one-shot Prisoner's Dilemma game

The (one-shot) Prisoner's Dilemma game involves two players each with two strategies: C (coöperate) and D (defect). In strategic form the game is:

	C	D
C	R	S
D	T	P

with the requirement that $T > R > P > S$ and that $2 \cdot R > T + S$. Typically, even usually, in experiments $T = 5$, $R = 3$, $P = 1$, and $S = 0$. Since $T > R$ and $P > S$, there is only one equilibrium point (EP): both players play D . The dilemma, of course, is that if both players could play C , both would be better off, since $R > P$.

Game theory says the teacher is wrong

John Nash proved that, allowing pure and mixed strategies, every finite n -person game has at least one equilibrium point (EP).

The *Nash equilibrium solution concept* holds that the “solution” or predicted outcome of any game (among rational players) will be an EP. Since the one-shot prisoner’s dilemma has only one EP, the Nash equilibrium solution concept predicts that both players will defect.

Further, for any fixed number, n , of iterations of Prisoner’s Dilemma, a backwards induction argument suffices to prove that the n -period Iterated Prisoner’s Dilemma (IPD or RPD) game still has exactly one EP: both players play D in each game.

Experiments say the teacher is right

It is interesting, and significant, that in the first human experiment with repeated prisoner's dilemma the human subjects were asked to record their thoughts as the game was being played. Comments such as

- “Perverse!”
- “Oh ho! Guess I’ll have to give him another chance.”
- “In time he could learn, but not in ten moves so:”
- “What’s he doing?!!”

- “I’m completely confused. Is he trying to convey information to me?”
and
- “This is like toilet training a child—you have to be very patient.”

appear throughout the 100 iterations of the game. Even so, the two subjects jointly cooperated in 60 of the 100 iterations. By the lights of classical game theory this was a remarkably rewarding triumph of irrational behavior.¹

¹These results are not inconsistent with subsequent empirical findings.

Consider the one-shot PD

By defecting, the teacher can guarantee that she will receive a payoff greater than S . Similarly, by not announcing a surprise exam as she did, the teacher can avoid uttering a falsehood (in that case). If the teacher/player is willing to undertake some risk, however, there is also a chance that the teacher can do better than to receive P or to do without a surprise exam. Suppose the teacher has the following policy: play C with probability r and play D with probability $(1 - r)$. If the other player plays the same strategy, then the expected return for each player, $E(r)$, is:

$$R \cdot r^2 + T \cdot (r - r^2) + P \cdot (1 - r)^2 \quad (1)$$

on the harmless assumption that $S = 0$. Rearranging, the gamble pays off

(in expectation) if

$$\frac{R \cdot r^2 + T \cdot (r - r^2)}{(1 - (1 - r)^2)} > P \quad (2)$$

or

$$\frac{R \cdot r + T \cdot (1 - r)}{(2 - r)} > P \quad (3)$$

Obviously, setting $r = 1$ (for both players) yields an expected value of R , and for fixed T , R , and P (with $S = 0$) this maximizes the expected value.

More interestingly, note that on the left-hand side both the numerator and the denominator are positive, so fixing r and P , it is always possible to increase R and T sufficiently to make the inequality hold. Of course, on the standard assumptions of game theory, this should not matter.

This is but a crude model

of how a player might reasonably deal with risk in the PD game. Still it tells us something. Let $d = T - R$. Prisoner's Dilemma requires that $0 < d < R$. Fix d at some small value, say $d = 2$ in:

$$\frac{R \cdot r + (R + d) \cdot (1 - r)}{(2 - r)} > P \quad (4)$$

In the usual PD problem, $R = 3, T = R + d = 3 + 2 = 5, P = 1, S = 0$. Using the $R = R + d$ formulation, fix d, P, S . We see from expression (4) that $E(r)$ (the expected return for each player, assuming an independent probability of r of cooperating) increases as R increases, for any fixed value of r . Suppose the outcome values are dollar amounts. Let $R = 100$ ($R = \$100, T = \102 etc.).

Consider now our rational human teacher

and her rational human counter-player. Each sees the situation; each understands that the goal is to maximize dollars captured individually, not to get more dollars than the other player. It remains true that each player can with certainty avoid the sucker's payoff S by defecting. . . . Surely, many players would reason that jointly they have much to lose by not cooperating and little extra to gain by individually defecting. Given that the other player has a coinciding interest in mutual cooperation, why not take a chance and play C ? Surely as well, the strength of such sentiments increases with R . Suppose $R = \$1,000,000$. Suppose it is much larger than that.

Believe it? OK. Don't? OK.

Consider now the teacher's 100-play IPD game. Even if the teacher and her counter-player both find the one-shot PD reasoning above unconvincing, they surely would give pause when faced with 100 plays, each with a reward of \$1,000,000 for mutual coöperation. Do they both really want to follow a policy of defecting every time, no matter what? Must they conclude that their actions cannot have an effect—positive or negative—on the other player? Surely that is an awfully strong and unduely pessimistic assumption about a supposedly rational player. Why not try coöperating and if it is reciprocated continue to do so?

If these numbers aren't convincing for these arguments, increase the number of plays to a million and R to a billion. Increase them all you want. If the teacher's r is large enough given T, R, P, S , the chance she will get her points is, I think we should agree, an excellent one.

The point may be summarized in the following manner

Suppose each player reasons that there is some chance, r , that the counter-player can be induced during the 100 iterations to play C fairly often. Given the counter-player's interest in maximizing returns (as opposed to gaining relative points only), is it reasonable to assume, without probing, that $r = 0$? Let there be exactly n iterations of the game, known to players. Let $P = \frac{1}{n}$. Let $d = T - R$ be small, as above. Are there no large values of n and R for which it would not be folly not to probe the counter-player for mutual cooperation? This obviously rhetorical question answers itself in the negative.

Comments

- “Birds do it, bees do it. . . .”
- People do it.
- Horses do it.

The Nash equilibrium is a solution *concept* for n-person games. The concept is that the games are solved by finding an EP (equilibrium point). That is where the players will end up. Long-established experience has shown that humans in (definitely) Iterated Prisoner’s Dilemma games consistently do better than mutual all defect. Assuming no pervasive flaws in experiments conducted over a 50-year period, either the human subjects are consistently, egregiously irrational, or the Nash equilibrium concept is flawed (or both).

Contra Nash

The fundamental error on the part of the students was to assume that the teacher's r is, or even must be, 0. The students assumed that in making the tradeoff between risk (of speaking falsely) and reward (being able to offer a surprise exam), the teacher would place no value (or no sufficiently large value) on the reward, at the expense of taking some risk. Similarly, I have argued, in the definitely IPD both players face a tradeoff between risk (of getting the sucker's payoff, S) and reward (achieving P very often during the iterations). The flaw in the Nash equilibrium solution concept (at least for IPD) is to impose the assumption that both players are unwilling to trade any risk at all (however small) for any reward at all (however large).

But Where's the Fallacy?

In the Surprise Exam paradox the induction can't get started. Where exactly is the misstep in the backwards induction argument in the IPD case? Recall: Let $G(i)$ be the situation with i games left to play. To make an induction argument, we need to show two things:

1. Base step: If $G(1)$, then Condition. FSA: Let's grant this for IPD.
2. Induction step: If Condition at $G(n)$, then Condition at $G(n + 1)$.

Think of modeling the reasoning as a dynamic programming problem. Let $p_i = \text{prob}(C_i)$ = the probability that column chooser will cooperate (play C) with i games to go.

But where's the fallacy?

The DP model assumes that p_i is independent of the history of play, of what happened at stages $N, N - 1, \dots, i + 1$. Letting o_i be the outcome of the play at stage i , the backwards induction argument or dynamic programming model assumes that

$$p_i = \text{prob}(C_i | o_N, o_{N-1}, \dots, o_{i+1}) = \text{prob}(C_i) \quad (5)$$

If this assumption is violated, then the backwards induction argument fails. In the lingo of dynamic programming, the *Markov property* (roughly, that it doesn't matter how you get to a given stage, i , only that you are there) does *not* obtain. And often it won't in fact.

Back to the proof by mathematical induction: the induction step ("if N , then $(N + 1)$ ") often in fact fails.

The fallacy

Because D in Prisoner's Dilemma strictly dominates C for either player, probabilities are thought not to matter. Taking now the row player's perspective we could introduce probabilities for the column player's plays as follows, with $p_i = \text{prob}(C_i) =$ the probability that column chooser will cooperate (play C) with i games to go (i.e., p_1 on the last game played, p_2 on the next to last, etc.).

	C	D
C	Rp_i	$S(1 - p_i)$
D	Tp_i	$P(1 - p_i)$

If p_i is fixed for all i , then ALL DEFECT is the best you can do. If, however, p_i is responsive to earlier play (cooperative, retaliatory), then it surely may be rational to do what you can to increase it.

Again on the Induction Step

Recall the one-shot PD:

	C	D
C	R	S
D	T	P

with the requirement that $T > R > P > S$ and that $2 \cdot R > T + S$.

Consider the one-shot PD in this form, Pattern 1

	C	D
C	R	S
D	T	P

 \implies

	C	D
C	B	S
D	$B + \epsilon$	$B - \epsilon$

($\epsilon > 0$, let $S = 0$, assume $B - \epsilon > S$)

- Note: $2 \cdot R > T + S$, but possibly $T + S > 2 \cdot P$.
- Fix B , make N (the number of plays) be smallish, make ϵ head towards 0. Are you willing to risk cooperation?

Now this version, Pattern 2

		C	D
C	R	S	T
D	T	P	P

 \implies

		C	D
C	B	S	$B + \epsilon$
D	$B + \epsilon$	$S + \epsilon$	$S + \epsilon$

$(\epsilon > 0, \text{ let } S = 0)$

- Let B and N increase arbitrarily, and ϵ decrease arbitrarily. Are there no values at which you would risk cooperation?

Summing up on Definitely IPD

- To succeed, the backwards induction argument for DIPD must recommend *ALL DEFECT* in both patterns (above), regards of parameter values.
- It is plausible and not unreasonable for the teacher to have a risk/return tradeoff allowing her to offer a surprise exam.
- It is plausible and not unreasonable for both players in DIPD to have risk/return tradeoffs allowing them to try some cooperation.
- Think of *TIT FOR TAT* as a simple reinforcement schedule.

A Game Theorist's Counter, with Response

C: Nothing new here. Of course if your opponent is irrational it may be rational to try some cooperation. That's old news.

R: Name calling merely evades the issue. You can insist on your definition of rationality, but you can't thereby make the concept interesting, descriptively sound, broadly useful, and free of paradox.

C: What positive concept do you have, that is more interesting, descriptively sound, broadly useful, and free of paradox?

R: A fair question.

A beginning of an answer

- *adaptive (learning) agent* \approx responds to experience and information, and modifies its behavior. Recognized in the literature.
- *exploring agent* \approx takes risks to obtain experience and information (and then is adaptive).
- Metaheuristics (metastrategies), procedures for finding solutions (strategies).

A beginning of an answer (con't.)

Metaheuristics (metastrategies):

1. Local search (incestuous) methods, e.g., hill climbing, simulated annealing, Q-learning
2. Population-based (promiscuous) methods, e.g., Learning Classifier Systems, memetic algorithms, evolutionary computation

All: not merely adaptive, but also exploring.

Explosion of attention and innovation. They are proving interesting, broadly useful, often descriptively sound.

And where are the paradoxes?

Counter and Reply, again

C: But can't this be incorporated into a Bayesian framework, so that we're back where we started?

R: In principle, yes, but this requires huge computational power. Also:

1. There is a premium on simpler explanations. For 'lower' species (animals and plants and bacteria, etc.). To understand just how far we can get with simpler models.
2. Even if we have (super) humans with big computers, the computational power needed is often way beyond reach. Beyond intractability we even have undecidability.

Counter and Reply, and again

C: Then haven't we just discovered bounded rationality again?

R: Yes, and the *Mona Lisa* is just oil paint on canvas.

1. Exploring rationality may be seen as a special kind of bounded rationality. Compare with satisficing.
2. Exploring rationality is appropriate even with essentially unlimited computational bounds, due to intractability and undecidability.
3. Nothing in the arguments I gave required bounded rationality. Riskless rationality leads to paradoxes and Pareto-inferior solutions. A risky—or *exploring*—rationality will often serve us better even when unbounded. What is true in the finite case need not be false in the infinite case.

The upshot

- In the Surprise Exam, the students failed to recognize that the teacher faces a risk/return tradeoff. The teacher can give a surprise exam if she is willing to undertake some risk of speaking falsely.
- Like the students, classical, game-theoretic (riskless) rationality takes an extreme, limiting position on the risk/return tradeoff. It may be wise (and rational) to explore with a counter-player at the risk of some loss.
- This leads to a notion of an exploring rationality, for which there is a rich and fruitful body of algorithms instantiating the concept.
- Agent-based modeling is a natural and entirely appropriate tool for exploring exploring rationality.

Extra Foils

In case there's time.

To handle in discussion if it comes up: risk versus uncertainty. I think my arguments work in both cases.

On Beyond Nash (Something positive)

Suppose we have a game with n players (in IPD, this n is 2). At the conclusion of play, each player i has played some strategy, s_i . This constitutes the *strategic configuration* or SC .

$$SC = (s_1, s_2, \dots, s_n) \quad (6)$$

Let $H_i(s_1, s_2, \dots, s_n) = H_i(SC)$ be the payoff to player i given the strategies, including i 's, played in SC . An *equilibrium point* or EP is any SC^* such that for each $i = 1, 2, \dots, n$

$$H_i(SC^*) = \max_{s_i} H_i(s_1^*, s_2^*, \dots, s_i, \dots, s_n^*) \quad (7)$$

In other words, an *EP* (equilibrium point) is an *SC* (strategic configuration) such that no individual player can (or could) unilaterally do better by picking a different strategy than the one the player has in *SC*. The Nash equilibrium solution concept for games is that among rational players every game will conclude at an *EP*.

Generalizing the Nash solution concept

Given s , a particular SC , define the *improvement vector* for s or IV_s , as

$$IV_s = (c_1, c_2, \dots, c_n) \quad (8)$$

where c_i is the *count* or number of ways i distinct players can jointly alter their strategies in such a way that each of the i players does equally well or better. More carefully, if $c_3 = 2$ then there are two distinct groups of 3 players who collectively have strategies that if taken, while all the strategies outside the group remain as in s the relevant SC , would make everyone in the group no worse off. Now a single group of 3 (or whatever) might have many ways to do this. IV_s ignores this and only counts the number of groups of a given size with 1 or more opportunities. (We can also define a *strict improvement vector* or SIV , as an IV in which everyone in every group is strictly better off.)

Points arising

1. The IV concept is a generalization of the EP concept. $IV_{SC} = (0, c_2, \dots, c_n)$ if and only if SC is an EP .
2. The payoff vector, H_s , of a strategic configuration s is denoted:

$$H_s = (H_1(s_1), H_2(s_2), \dots, H_n(s_n))$$

3. In the one-shot PD, the *table of space of improve vectors* or the *improvement space table* or IS is:

SC	IV	H_{SC}	target Hs
(D, D)	$(0, 1)$	(P, P)	(R, R)
(D, C)	$(1, 0)$	(T, S)	(P, P)
(C, D)	$(1, 0)$	(S, T)	(P, P)
(C, C)	$(2, 0)$	(R, R)	(T, S) (S, T)

where the target Hs are the payoffs if the associated IVs are realized.

4. We might similarly define how the players might act to make things worse. This would result in a disimprovement space table.
5. We say that SC_x *strictly dominates* SC_y iff for all $i = 1, 2, \dots, n$,

$$H_i(s_i^x) > H_i(s_i^y)$$

or equivalently

$$H_x > H_y$$

Note that (C, C) strictly dominates (D, D) in the one-shot PD. (Hence the dilemma for the Nash solution concept.)

6. In the n-shot IPD, there is a correspondingly larger improvement space, generated by the Cartesian product of the one-shot spaces.
7. Any well-defined game will have well-defined improvement and disimprovement spaces. In learning to play the game, or in rational deliberation about it, players should be thought of as exploring these spaces.

At risk of being wrong, but with potential for reward, players can form hypotheses about their counterparts, as play unfolds (actually or prospectively).

8. Given a well-defined game, with its improvement/disimprovement spaces and attendant payoff vectors, how play actually unfolds will depend upon the strategy formation and selection algorithms employed by the players.
9. In general and under realistic assumptions regarding strategy formation and selection algorithms, ending at an EP will be unusual in games with IVs having non-zero terms past the first position with attendant large payoffs.

Note: This is what the now large literature on computational dynamics in games has yielded—well something consistent with this.

10. We can look for and expect robustness results for classes of learning algorithms in games. Think: replicator dynamics, reinforcement learning. Convergence (and speed of it) to SCs with particular properties. Example: in Q-learning, with potential reward R and risk r , the players will achieve the R with such-and-such characteristics.

And that is how we can improve predictions and better understand decision processes (human or not).

Note vis à vis Evolutionary Game Theory

- Views here accord in many ways with evolutionary game theory, e.g., Gintis in *Game Theory Evolving*
- And also differ on the underlying theory of rationality. EGT wishes to save the classical theory (and the Nash equilibrium) by finding refinements (via evolutionary processes). How? We do what we do because of the way we're built.
- Yes, but. . . Why are we built the way we are?
Why has evolution produced the traits we have? Why are they adaptive (and rational in an extended sense)?
- Suggestion here: Rationality, properly understood, prompts us to be willing to take risks in strategic contexts. It's true that we are cognitively bound, but that's not—at least superficially—why we do better than ALLD in repeated prisoner's dilemma.

What Next?

Conjecture 1. [SAGE: Smart Agents Go Efficient] *Under reasonable discount rates, reasonable learning regimes, and indefinite (or definite but sufficiently long) repeated play of a stage game, agents can stably get to Pareto optimal outcomes.*

This is a conjecture and cannot be proved, but it could be disproved (or narrowed) in certain cases and demonstrated in other particular cases. Interesting to see when it obtains and when not. Let's take a preliminary look.

NB smart \approx probes the counter-player, collects information, revises accordingly

IPD Again

Suppose you are playing IPD and are confident of having a reasonably long run of it. Consider the following learning strategy.

1. Pick a run length, L_r over which you will explore the strategy space. Here, let $L_r = 3$, so you have 8 possible (nonreactive strategies).
2. Pick an iteration length, $L_i =$ the number of iterations you will play each tested strategy before deciding how to exploit the information you get from your exploratory experiment. Here again, let $L_i = 3$.
3. In random order, play each strategy L_i times in succession, recording what happens.

4. Evaluate the results and choose a revised policy of play.

What if Your Counter-Player Plays Tit-for-Tat?

No.	Strategy	Reward Sequence	Reward
1	(000)(000)(000)	(TPP)(PPP)(PPP)	$T+8P$
2	(100)(100)(100)	(RTP)(STP)(STP)	$3T+R+3P+2S$
3	(010)(010)(010)	(TST)(PST)(PST)	$4T+2P+3S$
4	(001)(001)(001)	(TPS)(TPS)(TPS)	$3T+3P+3S$
5	(110)(110)(110)	(RRT)(SRT)(SRT)	$3T+4R+2S$
6	(101)(101)(101)	(RTS)(STS)(RTS)	$3T+2R+4S$
7	(011)(011)(011)	(TSR)(TSR)(TSR)	$3T+3R+3S$
8	(111)(111)(111)	(RRR)(RRR)(RRR)	$9R$

Evaluation

Recall: $T > R > P > S$ and $2R > T + S$.

Well, fixing R, P, S , it's always possible to increase T so that $v(8) < v(i), i \neq 8$. OK, but $v(5) > v(2), v(4), v(6), v(7)$ What about $v(5) = v(110)$?

When is it that $v(8) = 9R > v(5) = 3T + 4R + 2S$? Assume $S = 0$. Then:

$$9R > 3T + 4R \text{ or } R > \frac{3}{5}T$$

Note: Let $L_i = n$, then $v(8) = n3R$ and $v(5) = nT + nR + R$ (assuming $S = 0$)

$$n3R > nT + nR + R \text{ when } R > \frac{n}{(2n-1)}T$$

$$\text{What about } v(3) = v(010)? \quad R > \frac{T(n+1)+P(n-1)}{3n}$$

$$\lim_{n \rightarrow \infty} \frac{T(n+1)+P(n-1)}{3n} = \frac{T+P}{3}$$

Discussion

So, if R is at all close to T , then the learner with this very simple regime will quickly learn to cooperate with the TIT-FOR-TAT player.

Also, for any values of R and T , the lesson can be learned with sufficiently high L_i .

What does it cost the TIT-FOR-TAT player to teach the lesson? When will it be worth it?

This is a very limited test and confirmation of the SAGE conjecture. What could we do to broaden and deepen our testing of it? What happens if we, say, try all $m-1$ strategies and both players are probing and learning?

Recall: smart \approx probes the counter-player, collects information, revises accordingly. What can we say about optimality and smartness?

\$Id: surprise-exam-foils.tex,v 1.7 2003/03/16 14:21:51 sok Exp \$