Incentive Compatibility Implies Signed Covariance*

Michael Suk-Young Chwe

Department of Political Science, UCLA

Los Angeles, CA 90095–1472

michael@chwe.net www.chwe.net

July 2006

Abstract

When a person's payoff depends on both her action and probabilistic events, the action she chooses and her payoff can be understood as random variables. This paper shows that incentive compatibility implies that when a person chooses among two actions, conditional on these two actions, her action is nonnegatively correlated with the payoff difference between the two actions. This simple and robust result has implications in a wide variety of contexts, including individual choice under uncertainty, strategic form games, and incomplete information games. Incentive compatibility constraints have an immediate "statistical" interpretation.

JEL classification: C72 Noncooperative Games

Keywords: incentive compatibility, revealed preference, statistical game theory, correlated equilibrium, identification, incomplete information games, spatial data analysis, local interaction games

* I am grateful for comments and suggestions received at the 2005 Decentralization Conference at the University of Illinois, Champaign-Urbana, the University of California, Berkeley, the University of California, Irvine, the University of California, Los Angeles, the University of California, San Diego, the University of Chicago, Northwestern University, Rice University, the University of Texas, Austin, the University of Washington, and from anonymous referees. Conversations with Philip Bond and Jeff Lewis and a comment by Matt Jackson were particularly helpful. I also appreciate Yafang Su, of Academic Technology Services at UCLA, for her help in geocoding the flag data.

Introduction

If a person takes an action, she cannot prefer to take another action instead. This fundamental idea is called "revealed preference" in choice theory or "incentive compatibility" in game theory. Often a person's payoff, and hence her chosen action, depends on probabilistic events she cannot control, such as the weather or the actions of others. Thus her action, and the payoff she receives, can be considered random variables. This paper shows that incentive compatibility implies that when a person chooses among two actions, conditional on these two actions, her action is nonnegatively correlated with the payoff difference between the two actions. This simple result has implications in a wide variety of contexts, including individual choice under uncertainty, strategic form games, and incomplete information games. This result is shown in a general context, independent of any assumption about prior beliefs, communication and coordination devices, incomplete information, or whether the randomness is due to exogenous events or the endogenous actions of other people or both. This result shows how revealed preference or incentive compatibility constraints have an immediate "statistical" interpretation.

In the context of strategic form games, this result has several implications. For example, in a two person game in which one person's payoffs are quadratic, one can predict the sign of the covariance of people's actions in any correlated equilibrium of the game. In a local interaction game, one can predict the sign of the covariance between a person's action and the number of neighbors who take the same action. For 2×2 games, observing a signed covariance in people's actions is sufficient to identify pure strategy Nash equilibria of the game. For incomplete information games, the result yields testable empirical predictions which hold for any assumption about prior beliefs or the kind of incomplete information and which do not require computing an equilibrium. This paper starts with definitions and the main result, considers several examples, and concludes by discussing the merits of "statistical game theory."

Definitions and main result

We have a standard framework. A person chooses x from a finite set X but does not choose y, which belongs to a finite set Y. For example, y might be determined by exogenous randomness, the choices of other people who have their own motivations, or both. Her utility function is given by $u : X \times Y \to \Re$. Let $p : X \times Y \to \Re$ be a probability distribution over

$X \times Y$, in other words $p(x, y) \geq 0$ for all $(x, y) \in X \times Y$ and $\sum_{(x,y) \in X \times Y} p(x, y) = 1$. Let U be the set of all utility functions on $X \times Y$ and let P be the set of all probability distributions on $X \times Y$.

Incentive compatibility is defined by the following linear inequality.

$$\sum_{y \in Y} p(x, y)u(x, y) \geq \sum_{y \in Y} p(x, y)u(x', y) \text{ for all } x, x' \in X. \qquad (IC)$$

The idea here is that when the person plays x, the probability distribution over Y is given by $p(x, y)$, and her expected utility is $\sum_{y \in Y} p(x, y)u(x, y)$. If she plays x' instead, then her payoffs change but the resulting probability distribution over Y does not change (since she cannot control y), and hence she gets expected utility $\sum_{y \in Y} p(x, y)u(x', y)$. The IC constraint says that she cannot gain by doing so. Let $IC(u)$ be the set of probability distributions p which satisfy IC given u, and let $IC(p)$ be the set of utility functions u which satisfy IC given p. It is easy to see that $IC(u)$ and $IC(p)$ are convex sets. It is easy to see that $\bar{u} \in IC(p)$, where $\bar{u}$ is defined as $\bar{u}(x, y) = 0$ for all x, y, and hence $IC(p) \neq \emptyset$. We know that $IC(u) \neq \emptyset$ by for example Hart and Schmeidler (1989) and Nau and McCardle (1990). We say that u is trivial if $IC(u) = P$, in other words, if every probability distribution satisfies IC.

We can think of $X \times Y$ as a probability space with probability distribution p. A random variable is a function defined on $X \times Y$. Given some subset $Z \subset X \times Y$, we define the "indicator function" $\mathbf{1}_Z : X \times Y \to \Re$ as $\mathbf{1}_Z(x, y) = 1$ if $(x, y) \in Z$ and $\mathbf{1}_Z(x, y) = 0$ otherwise. If $x \in X$, for convenience we write $\mathbf{1}_x$ instead of $\mathbf{1}_{\{x\} \times Y}$, and similarly if $y \in Y$, we write $\mathbf{1}_y$ instead of $\mathbf{1}_{X \times \{y\}}$. We define $\mathbf{x} : X \times Y \to X$ as $\mathbf{x}(x, y) = x$ and $\mathbf{y} : X \times Y \to Y$ as $\mathbf{y}(x, y) = y$. In this paper, we use boldface to indicate random variables.

Given p and a real-valued random variable $\mathbf{f} : X \times Y \to \Re$, the expectation of $\mathbf{f}$ is $E_p(\mathbf{f}) = \sum_{(x,y) \in X \times Y} p(x, y)\mathbf{f}(x, y)$. Given $Z \subset X \times Y$, we write $p(Z) = \sum_{(x,y) \in Z} p(x, y)$. Given $Z \subset X \times Y$ such that $p(Z) > 0$, the conditional expectation of $\mathbf{f}$ is $E_p(\mathbf{f}|Z) = (\sum_{(x,y) \in Z} p(x, y)\mathbf{f}(x, y))/p(Z)$. The covariance of two real-valued random variables $\mathbf{f}$ and $\mathbf{g}$ is $cov_p(\mathbf{f}, \mathbf{g}) = E_p(\mathbf{fg}) - E_p(\mathbf{f})E_p(\mathbf{g})$, and the conditional covariance $cov_p(\mathbf{f}, \mathbf{g}|Z)$ is defined similarly. For convenience, if $p(Z) = 0$, we write $cov_p(\mathbf{f}, \mathbf{g}|Z) = 0$. Note that for random variables $\mathbf{f}, \mathbf{g}, \mathbf{g}'$ and real numbers $\alpha, \alpha' \in \Re$, we have $cov_p(\mathbf{f}, \alpha\mathbf{g} + \alpha'\mathbf{g}') = \alpha cov_p(\mathbf{f}, \mathbf{g}) + \alpha' cov_p(\mathbf{f}, \mathbf{g}')$.

The probability distribution p can be understood in a few different ways. Since the person controls $x \in X$ but not $y \in Y$, one might instead think of the person as choosing a conditional probability distribution $r(x|y)$, where y has an exogenous probability distribution $q(y)$ over Y. We use $p(x, y)$ because it is simpler and mathematically equivalent (although it is not exactly equivalent, since one might say that $r(x|y)$ should be well-defined even when $q(y) = 0$). Another interpretation is that p is the result of a possibly quite complicated messaging and information mechanism by which the person learns and updates beliefs about $y \in Y$; for example, the person might receive some signal which is correlated with y, she might receive an explicit message from another person about what y is, or she might know something about y directly. Regardless of how p occurs, if p violates IC then the person is not maximizing her expected payoff, since it is always possible for the person, whenever he chooses x, to choose x' instead (see for example Myerson 1991 on the "revelation principle"). Finally, p can be understood simply as the observed histogram of the person's actions over some time period. We might not know why the person's actions along with exogenous or endogenous randomness result in the histogram p, but we can surely say that whenever she chose x, she could have chosen x' instead, and thus she could not have gained by doing so. The IC constraints can thus be understood as revealed preference inequalities.

Our main result is a signed conditional covariance. Given incentive compatibility, then conditional on two choices x and x', the random variable which indicates when the person plays x and the random variable which is the payoff difference between x and x' are nonnegatively correlated.

Proposition. Say p, u satisfy IC and $x, x' \in X$. Then

$$cov_p(\mathbf{1}_x, u(x, \mathbf{y}) - u(x', \mathbf{y}) | \{x, x'\} \times Y) \geq 0.$$

This result is obtained by manipulating two IC constraints: the constraint that when the person chooses x, she cannot do better by playing x', and when the person chooses x', she cannot do better by playing x. The proof is in the appendix. Note that the IC constraints are linear in p while the covariance in the Proposition is quadratic in p; in other words, the Proposition is not a linear restatement of the IC constraints.

Since the Proposition is based on only the IC constraints, it holds under very weak conditions. Again, the IC constraints hold regardless of any assumption about how the person might or might not have knowledge about y. The IC constraints do not involve any assumption about how exactly $p(x, y)$ comes about. The IC inequalities are minimal requirements for individual rationality: whenever the person chooses x, she could have chosen x' instead, and the fact that she did not means that she could not have gained by doing so.

The 2×2 *case*

To illustrate the Proposition, we take the simplest nontrivial case, when $X = \{x, x'\}$ and $Y = \{y, y'\}$. Say that $u : X \times Y \to \Re$ is given by the following table.

	y	y'
x	8	3
x'	0	7

The Proposition says that if p, u satisfy IC, then the two random variables $\mathbf{1}_x$ and $u(x, \mathbf{y}) - u(x', \mathbf{y})$ are nonnegatively correlated given p. These two random variables are shown below.

	y	y'
x	1	1
x'	0	0

$\mathbf{1}_x$

	y	y'
x	8	-4
x'	8	-4

$u(x, \mathbf{y}) - u(x', \mathbf{y})$

For these two random variables to have nonnegative covariance, it must be that when $\mathbf{1}_x$ is high, $u(x, \mathbf{y}) - u(x', \mathbf{y})$ is high; roughly speaking, it must be that $p(x, y)$ and $p(x', y')$ are large compared with $p(x, y')$ and $p(x', y)$.

Here are some example probability distributions.

	y	y'
x	0.6	0
x'	0	0.4

p

	y	y'
x	0.55	0.04
x'	0.05	0.36

p'

	y	y'
x	0.7	0.3
x'	0	0

p''

	y	y'
x	0.1	0.3
x'	0.5	0.2

p'''

The first distribution p is what results if y occurs with probability 0.6 and y' occurs with probability 0.4, and the person knows exactly when either y or y' occurs, and makes his optimal choice accordingly. The second distribution p' is consistent with a situation in which the person gets a noisy signal about whether y or y' occurs; the signal is correct often enough so that the person still chooses x if the signal indicates y and x' if the signal indicates y'. The third distribution p'' is consistent with the person not knowing anything about y or y'; since y is more likely than y', the person is best off choosing x all the time.

All three distributions p, p', p'' satisfy IC, and in all three distributions, the covariance of $\mathbf{1}_x$ and $u(x, \mathbf{y}) - u(x', \mathbf{y})$ is nonnegative (positive in p and p' and zero in p''). In the last distribution p''', the covariance of $\mathbf{1}_x$ and $u(x, \mathbf{y}) - u(x', \mathbf{y})$ is negative, and it is easy to see that IC is violated: regardless of the beliefs behind the person's choice and what he knows about y or y', he violates rationality because all the times that he chooses x and gets expected utility $(0.1)8 + (0.3)3 = 1.7$, he could choose x' instead and get a higher expected utility $(0.1)0 + (0.3)7 = 2.1$.

In the 2×2 case, we have two simple facts.

Fact 1. Say $X = \{x, x'\}$ and $Y = \{y, y'\}$ and u is nontrivial. Then either $cov_p(\mathbf{1}_x, \mathbf{1}_y) \geq 0$ for all $p \in IC(u)$ or $cov_p(\mathbf{1}_x, \mathbf{1}_y) \leq 0$ for all $p \in IC(u)$.

In other words, as long as u is nontrivial (that is, $P \neq IC(u)$), then we can sign the covariance of x and y in all incentive compatible p. Fact 1, proved in the appendix, follows immediately from the Proposition: the Proposition signs the covariance of $\mathbf{1}_x$ and $u(x, \mathbf{y}) - u(x', \mathbf{y})$, but when Y has only two elements, $u(x, \mathbf{y}) - u(x', \mathbf{y})$ is a linear function of $\mathbf{1}_y$ and hence we can sign the covariance of $\mathbf{1}_x$ and $\mathbf{1}_y$. Fact 2 follows immediately from Fact 1.

Fact 2. Say $X = \{x, x'\}$ and $Y = \{y, y'\}$ and u is nontrivial. Say $p, p' \in IC(u)$. If $cov_p(\mathbf{1}_x, \mathbf{1}_y) > 0$, then $cov_{p'}(\mathbf{1}_x, \mathbf{1}_y) \geq 0$. If $cov_p(\mathbf{1}_x, \mathbf{1}_y) < 0$, then $cov_{p'}(\mathbf{1}_x, \mathbf{1}_y) \leq 0$.

In other words, say that we observe p and then try to identify incentive compatible u. Once we have done this, it is natural to then make a prediction based on what has been learned about u. In the 2×2 case, this could not be simpler. If one observes for example positive covariance between $\mathbf{x}$ and $\mathbf{y}$, in any utility function u consistent with this observation, in any incentive compatible p' of any such utility function, one must have nonnegative covariance (assuming the utility function is not trivial). In other words, if we observe signed covariance, we can predict that in future behavior, the covariance cannot have the opposite sign. We can make this prediction without knowing or assuming anything else about the utility function (other than that it is nontrivial).

Quadratic payoffs

Say $X \subset \Re$ and $Y \subset \Re^m$; in other words, $x \in X$ is a real number and $y = (y_1, \ldots, y_m) \in Y$ is a vector of real numbers. When the utility function u is quadratic in x and y_i (actually, when it satisfies a somewhat weaker condition), we can sign the weighted sum of covariances between $\mathbf{x}$ and the various $\mathbf{y}_i$. The proof of Fact 3 is in the appendix.

Fact 3. Say that $X \subset \Re$, $Y \subset \Re^m$ and u satisfies the condition that $u(x, y) - u(x', y) = v(x, x') \sum_{j=1}^{m} c_j y_j + w(x, x')$, where $c_j \in \Re$ and $v(x, x') > 0$ when $x > x'$. Say p, u satisfy *IC*. Then $\sum_{j=1}^{m} c_j \, cov_p(\mathbf{x}, \mathbf{y}_j) \geq 0$.

For example, say that $X, Y \subset \Re$ and u is quadratic: $u(x, y) = k_{xy}xy + k_{xx}x^2 + k_{yy}y^2 + k_x x + k_y y + k_0$, where $k_{xy}, k_{xx}, k_{yy}, k_x, k_y, k_0 \in \Re$; we can think of k_{xy} as the "interaction term." It is easy to see that this utility function satisfies the condition in Fact 3: set $c_1 = k_{xy}$, $v(x, x') = x - x'$ and $w(x, x') = k_{xx}(x^2 - (x')^2) + k_x(x - x')$. Fact 3 says that $k_{xy} \, cov_p(\mathbf{x}, \mathbf{y}) \geq 0$. Thus we have three conclusions. First, if $k_{xy} \neq 0$, we can predict the sign of the covariance of $\mathbf{x}$ and $\mathbf{y}$. Second, if we observe a positive covariance, we can conclude that $k_{xy} \geq 0$. Third, say that we observe a signed covariance. We can predict that in any utility function consistent with this observation, in any behavior consistent with any such utility function, the covariance must have the same sign. We make this prediction knowing nothing else about the utility function, other than assuming it is quadratic and $k_{xy} \neq 0$.

It is easy to explain how the proof works. The Proposition signs the covariance between $\mathbf{1}_x$ and the utility difference $u(x, \mathbf{y}) - u(x', \mathbf{y})$ conditional on $\{x, x'\} \times Y$. When the condition in Fact 3 is satisfied, the utility difference $u(x, \mathbf{y}) - u(x', \mathbf{y})$ is a weighted sum of the various $\mathbf{y}_j$, and thus we can sign the weighted sum of covariances between $\mathbf{1}_x$ and the various $\mathbf{y}_j$ conditional on $\{x, x'\} \times Y$. It is not hard to show (Lemma 3 in the appendix) that the unconditional covariance $cov_p(\mathbf{x}, \mathbf{y}_j)$ can be obtained by summing up the conditional covariances $cov_p(\mathbf{x}, \mathbf{y}_j | \{x, x'\} \times Y)$ over $x, x' \in X$. Hence we can sign the weighted sum of unconditional covariances.

For an example which is not quadratic, say $u(x, y_1, y_2) = 5x^{1/2}y_1 - 3x^{1/2}y_2 + 4y_1y_2 + 6(y_1 - y_2)^2 - 7x^{3/2}$. By Fact 3, we have $5cov_p(\mathbf{x}, \mathbf{y}_1) - 3cov_p(\mathbf{x}, \mathbf{y}_2) \geq 0$.

Odds ratios

We can also think in terms of "odds ratios." Note that the person does not control y and hence cannot determine the absolute levels of $p(x, y)$ and $p(x', y)$, for example. However, the person is in control of x, and can determine the odds ratio $p(x, y)/p(x', y)$. One might think that if p, u satisfy IC, then the ratio $p(x, y)/p(x', y)$ should be higher when the payoff difference $u(x, y) - u(x', y)$ is higher. This is not true: the ratio does not always increase in the payoff difference. However, we can make a weaker statement: Fact 4 says that the odds ratio cannot always decrease in the payoff difference.

Fact 4. Say that p, u satisfy IC and $x, x' \in X$. Say that $u(x, y) - u(x', y)$ is not constant in y and that $p(x', y) > 0$ for $y \in Y$. The following statement is not true:

$$u(x, y) - u(x', y) > u(x, y') - u(x', y') \Leftrightarrow p(x, y)/p(x', y) < p(x, y')/p(x', y').$$

The proof is in the appendix, but it is easy to explain how it works. By the Proposition, we know that the person's action and his payoff difference are nonnegatively correlated. If x is always played less often relative to x' when the payoff difference is higher, then we would have a negative correlation.

Say $X, Y \subset \Re$. If $u(x, y) - u(x', y)$ strictly increases in y for all $x > x'$, then u is called strictly supermodular. If $p(x, y)/p(x', y)$ strictly increases in y for all $x > x'$, then p is called strictly totally positive of order 2 (Karlin and Rinott 1980a; Milgrom and Weber 1982 use the term "affiliated"). If $p(x, y)/p(x', y)$ strictly decreases in y for all $x > x'$, then p is called strictly reverse rule of order 2 (Karlin and Rinott 1980b). From Fact 4, if payoffs u are strictly supermodular, then p is not necessarily strictly totally positive of order 2, but cannot be strictly reverse rule of order 2. Similarly, if p is strictly totally positive of order 2, then u need not be strictly supermodular, but $-u$ cannot be strictly supermodular.

Choice under uncertainty

To illustrate our results in the context of choice under uncertainty, consider a simple example of a paparazzo and a celebrity. The paparazzo is a photographer who wants to get as close as possible to the celebrity, whose location changes each day. Let $x \in X$ be the photographer's location and $y \in Y$ be the celebrity's location, where $X, Y \subset \Re$. An appropriate utility function for the paparazzo is $u(x, y) = -(x - y)^2$.

If the paparazzo always knows the celebrity's location y, he chooses $x = y$, which maximizes $u(x, y)$. But the celebrity might wear disguises, the paparazzo might try to cultivate informants, the celebrity might try to create false rumors, and so forth, in a rather complicated process. The standard way to predict which location the paparazzo chooses on a given day, however, is quite straightforward: we simply specify the paparazzo's prior belief over Y of the celebrity's location on that day, and then we find the $x \in X$ which maximizes the paparazzo's expected payoff given this belief.

This paper's approach does not require specifying any beliefs. We write $u(x, y) = -x^2 + 2xy - y^2$. Since the coefficient on the xy term is 2, by Fact 3 we have $2cov_p(\mathbf{x}, \mathbf{y}) \geq 0$ and hence $cov_p(\mathbf{x}, \mathbf{y}) \geq 0$. In other words, we predict that the paparazzo's location and the celebrity's location are nonnegatively correlated. If the paparazzo always knows the celebrity's exact location, then $x = y$ and their locations are perfectly positively correlated. If the paparazzo never knows anything about the celebrity's location and thus always goes to the middle of town, their locations have zero correlation. What cannot happen is for the paparazzo's location and the celebrity's location to be negatively correlated. This prediction holds for any prior belief, for any specification of how the paparazzo gains information or disinformation, and for any probability distribution of the celebrity's actual location. This prediction holds regardless of the definition of $X \subset \Re$ and $Y \subset \Re$, for example whether the elements of X and Y are bunched together or evenly spread out. This prediction holds regardless of whether the celebrity consciously chooses his location or whether his location is determined exogenously by his shooting schedule. This prediction holds regardless of the celebrity's motivations, whether the celebrity despises and actively avoids the paparazzo, is indifferent or unaware, or in fact enjoys having his picture taken. This prediction is based only on the paparazzo's utility function and the incentive compatibility constraints.

Note that by Fact 3, this prediction holds even if the paparazzo's utility function has the form $u(x, y) = v(x) + w(y) - (x - y)^2$, where $v(x)$ is a function only of x and $w(y)$ is a function only of y. For example, the paparazzo might prefer locations in his own neighborhood, or prefer that the celebrity be in a location with good natural light. The essential aspect of the utility function which drives the result is the negative coefficient -2 on the xy term.

Now make the situation slightly more complicated. Say the celebrity's talent agency hires a security thug to harass paparazzi. As before, the paparazzo chooses a location $x \in X \subset \Re$, but now has to think about the celebrity's location y_1 and the security thug's location

y_2, where $y = (y_1, y_2) \in Y \subset \Re^2$. Say that the paparazzo's utility function is $u(x, y) = -(x - y_1)^2 + k(x - y_2)^2$, where $k \geq 0$ is a parameter indicating how much the paparazzo worries about the security thug. The paparazzo wants to be close to the celebrity but far from the security thug. We can write $u(x, y) = -x^2 + 2xy_1 - y_1^2 + kx^2 - 2kxy_2 + ky_2^2$. Hence by Fact 3 we know $2cov_p(\mathbf{x}, \mathbf{y}_1) - 2k\, cov_p(\mathbf{x}, \mathbf{y}_2) \geq 0$ and thus $cov_p(\mathbf{x}, \mathbf{y}_1) \geq k\, cov_p(\mathbf{x}, \mathbf{y}_2)$.

This prediction is easy to interpret. If the paparazzo pursues the celebrity and avoids the security thug, we have $cov_p(\mathbf{x}, \mathbf{y}_1) \geq 0$ and $cov_p(\mathbf{x}, \mathbf{y}_2) \leq 0$, and this is possible regardless of how large k is. It is impossible to have $cov_p(\mathbf{x}, \mathbf{y}_1) < 0$ and $cov_p(\mathbf{x}, \mathbf{y}_2) > 0$; it can never be that the paparazzo and celebrity negatively covary and the paparazzo and the security thug positively covary. If $cov_p(\mathbf{x}, \mathbf{y}_1)$ and $cov_p(\mathbf{x}, \mathbf{y}_2)$ are both positive, then $cov_p(\mathbf{x}, \mathbf{y}_1)$ must be sufficiently high; the more the paparazzo worries about the security thug, the higher k is, and the higher $cov_p(\mathbf{x}, \mathbf{y}_1)$ must be to make the harassment risk worthwhile. Finally, if $cov_p(\mathbf{x}, \mathbf{y}_1)$ and $cov_p(\mathbf{x}, \mathbf{y}_2)$ are both negative, then again $cov_p(\mathbf{x}, \mathbf{y}_1)$ must be sufficiently high. For small k, it cannot be that the paparazzo's covariance with the security thug is slightly negative but his covariance with the celebrity is very negative; by avoiding the security thug he cannot pay too high a price in terms of celebrity access. As k increases, the paparazzo tolerates greater avoidance of the celebrity in order to avoid the security thug.

Some special cases are interesting. If $cov_p(\mathbf{x}, \mathbf{y}_2) = 0$, then we have the same prediction as before, $cov_p(\mathbf{x}, \mathbf{y}_1) \geq 0$; if the paparazzo and security thug are uncorrelated, we can think of this as the security thug imposing a dead-weight cost on the paparazzo which is the same whatever the paparazzo does and therefore does not affect his decision. If the security thug and celebrity are always together, then $cov_p(\mathbf{x}, \mathbf{y}_1) = cov_p(\mathbf{x}, \mathbf{y}_2)$. Hence if $k < 1$, we must have $cov_p(\mathbf{x}, \mathbf{y}_1) \geq 0$; the security thug is not a sufficient deterrent. If $k > 1$, the security thug is scary enough and we have $cov_p(\mathbf{x}, \mathbf{y}_1) \leq 0$.

We can also identify k given observed behavior. As mentioned above, if $cov_p(\mathbf{x}, \mathbf{y}_1) \geq 0$ and $cov_p(\mathbf{x}, \mathbf{y}_2) \leq 0$, then there is no restriction on k. If we observe $cov_p(\mathbf{x}, \mathbf{y}_1) < 0$ and $cov_p(\mathbf{x}, \mathbf{y}_2) > 0$, this is impossible and thus we must reject our utility function. If $cov_p(\mathbf{x}, \mathbf{y}_1)$ and $cov_p(\mathbf{x}, \mathbf{y}_2)$ are both positive, we can conclude that $k \leq cov_p(\mathbf{x}, \mathbf{y}_1)/cov_p(\mathbf{x}, \mathbf{y}_2)$. If the paparazzo covaries weakly with the celebrity but strongly with the security thug, then k must be small; the paparazzo must not care much about the security thug since he doesn't mind covarying with him even for a low reward. If $cov_p(\mathbf{x}, \mathbf{y}_1)$ and $cov_p(\mathbf{x}, \mathbf{y}_2)$ are both negative, we conclude that $k \geq cov_p(\mathbf{x}, \mathbf{y}_1)/cov_p(\mathbf{x}, \mathbf{y}_2)$. If the paparazzo covaries strongly negatively with the celebrity and weakly negatively with the security thug, then k must be large, since the paparazzo pays a high price avoiding the celebrity just to slightly avoid the security thug.

Again, all of the conclusions here are robust, independent of any assumption about prior beliefs, the motivations of the celebrity and security thug and whether they choose consciously or not, whether they purposefully coordinate their locations or not, how the paparazzo learns about the celebrity and security thug, the definitions of X and Y, and so forth. The only assumptions here are the paparazzo's utility function itself and incentive compatibility. This robustness is nice for making predictions but especially desirable for identification. Our restrictions on k above hold under extremely weak assumptions and thus are almost unarguable. Our restrictions on k are also easy to compute and understand; finding restrictions on k using a more standard approach, involving assumptions about beliefs, utility "shocks" and error terms, would be much more complicated.

Strategic form games

In the standard finite strategic form game, we have a finite set of people $N = \{1, \ldots, n\}$, each with a finite strategy set A_i and each with a utility function $u_i : A \to \Re$, where $A = \times_{i \in N} A_i$. To put a strategic form game in our framework, for each player $i \in N$ we simply let $X = A_i$, $Y = A_{-i}$, and $u = u_i$, where $A_{-i} = \times_{j \in N \setminus \{i\}} A_j$. We thus have a set of constraints IC_i for each person i. A probability distribution p which satisfies IC_i for all $i \in N$ is called a correlated equilibrium (Aumann 1974). As is well known, any distribution over A resulting from a pure strategy or mixed strategy Nash equilibrium, or convex combination of Nash equilibria, is a correlated equilibrium. We define the random variable $\mathbf{a}_i : A \to A_i$ as $\mathbf{a}_i(a) = a_i$. Given a game u, we say $CE(u)$ is the set of correlated equilibria.

Consider the example 2×2 games below. In the first game, "chicken," given u_1 and the resulting IC constraints for person 1, we conclude using Fact 1 that $cov_p(\mathbf{a}_1, \mathbf{a}_2) \leq 0$. Given u_2 and the resulting IC constraints for person 2, we conclude using Fact 1 that $cov_p(\mathbf{a}_2, \mathbf{a}_1) \leq 0$. Hence we have $cov_p(\mathbf{a}_1, \mathbf{a}_2) \leq 0$ for all correlated equilibria p of this game. Similarly we conclude that $cov_p(\mathbf{a}_1, \mathbf{a}_2) \geq 0$ for all correlated equilibria p of the second game, "battle of the sexes."

$$
\begin{array}{ccc} & 0 & 1 \\ 0 & 3,3 & 1,4 \\ 1 & 4,1 & 0,0 \\ \multicolumn{3}{c}{cov_p \le 0} \end{array}
\qquad
\begin{array}{ccc} & 0 & 1 \\ 0 & 2,1 & 0,0 \\ 1 & 0,0 & 1,2 \\ \multicolumn{3}{c}{cov_p \ge 0} \end{array}
\qquad
\begin{array}{ccc} & 0 & 1 \\ 0 & 1,0 & 0,1 \\ 1 & 0,1 & 1,0 \\ \multicolumn{3}{c}{cov_p = 0} \end{array}
\qquad
\begin{array}{ccc} & 0 & 1 \\ 0 & 5,0 & 5,8 \\ 1 & 6,4 & 6,4 \\ \multicolumn{3}{c}{cov_p = 0} \end{array}
\qquad
\begin{array}{ccc} & 0 & 1 \\ 0 & 5,3 & 7,3 \\ 1 & 5,2 & 7,2 \\ \multicolumn{3}{c}{CE(u) = P} \end{array}
$$

In the third game, "matching pennies," given u_1 and the resulting IC constraints for person 1, we conclude using Fact 1 that $cov_p(\mathbf{a}_1, \mathbf{a}_2) \ge 0$. Given u_2 and the resulting IC constraints for person 2, we conclude using Fact 1 that $cov_p(\mathbf{a}_2, \mathbf{a}_1) \le 0$. Hence $cov_p(\mathbf{a}_1, \mathbf{a}_2) = 0$ for all correlated equilibria p of this game. In this game the unique correlated equilibrium is the mixed strategy Nash equilibrium, in which people's actions are chosen independently and hence have zero covariance. In the fourth game, the correlated equilibria are those distributions which place weight only on $(1,0)$ and $(1,1)$, since for person 1 action 0 is strongly dominated. Since there is no variation in person 1's action, the covariance between their actions is zero. In the fifth game, both u_1 and u_2 are trivial and $CE(u) = P$, that is, every distribution is a correlated equilibrium.

To identify 2×2 games, it turns out that a signed covariance is sufficient to identify a game's pure strategy Nash equilibria. In other words, if one observes a nonzero covariance between two people's actions in a 2×2 game, we need not care if their actions result from pure strategy Nash equilibria, mixed strategy Nash equilibria, correlated equilibria, or a mixture of all of these. The signed covariance itself is enough to locate pure Nash equilibria, knowing nothing else about the game. The proof of Fact 5 is in the appendix.

Fact 5. Say $n = 2$, $A_1 = \{a_1, b_1\}$, $A_2 = \{a_2, b_2\}$, and $p \in CE(u)$. If $cov_p(\mathbf{1}_{a_1}, \mathbf{1}_{a_2}) > 0$, then (a_1, a_2) and (b_1, b_2) are Nash equilibria of u. If $cov_p(\mathbf{1}_{a_1}, \mathbf{1}_{a_2}) < 0$, then (a_1, b_2) and (b_1, a_2) are Nash equilibria of u.

After identifying a 2×2 game to the extent possible given observations, what can we predict? From Fact 2, if we observe for example positive covariance in some correlated equilibrium, in any such game consistent with this observation, in any correlated equilibrium of any such game, we can predict nonnegative covariance (as long as the game is nontrivial). In other words, observing a signed covariance is enough to make predictions about future play in the game, without knowing anything else about the game or assuming anything other than nontriviality.

For an example of a three-person game, consider the following "Three Player Matching Pennies Game" (Moreno and Wooders 1998).

	0	1			0	1
0	$1, 1, -2$	$-1, -1, 2$		0	$-1, -1, 2$	$-1, -1, 2$
1	$-1, -1, 2$	$-1, -1, 2$		1	$-1, -1, 2$	$1, 1, -2$
	0				1	

The Proposition says that $cov(\mathbf{a}_1, u_1(1, \mathbf{a}_2, \mathbf{a}_3) - u_1(0, \mathbf{a}_2, \mathbf{a}_3)) \geq 0$ in any correlated equilibrium. In this game, $u_1(1, \mathbf{a}_2, \mathbf{a}_3) - u_1(0, \mathbf{a}_2, \mathbf{a}_3) = -2(1 - \mathbf{a}_2)(1 - \mathbf{a}_3) + 2\mathbf{a}_2\mathbf{a}_3 = -2 + 2\mathbf{a}_2 + 2\mathbf{a}_3$. Since -2 is a constant, we have $cov(\mathbf{a}_1, 2\mathbf{a}_2 + 2\mathbf{a}_3) \geq 0$ and thus $cov(\mathbf{a}_1, \mathbf{a}_2) + cov(\mathbf{a}_1, \mathbf{a}_3) \geq 0$. Similarly, we find that $cov(\mathbf{a}_1, \mathbf{a}_2) + cov(\mathbf{a}_2, \mathbf{a}_3) \geq 0$. We also know that $cov(\mathbf{a}_3, u_3(\mathbf{a}_1, \mathbf{a}_2, 1) - u_3(\mathbf{a}_1, \mathbf{a}_2, 0)) \geq 0$ and that $u_3(\mathbf{a}_1, \mathbf{a}_2, 1) - u_3(\mathbf{a}_1, \mathbf{a}_2, 0) = 4(1 - \mathbf{a}_1)(1 - \mathbf{a}_2) - 4\mathbf{a}_1\mathbf{a}_2 = 4 - 4\mathbf{a}_1 - 4\mathbf{a}_2$. Thus $cov(\mathbf{a}_1, \mathbf{a}_3) + cov(\mathbf{a}_2, \mathbf{a}_3) \leq 0$. So we have three inequalities on the three covariances $cov(\mathbf{a}_1, \mathbf{a}_2)$, $cov(\mathbf{a}_1, \mathbf{a}_3)$, and $cov(\mathbf{a}_2, \mathbf{a}_3)$. From these inequalities, we conclude that $cov(\mathbf{a}_1, \mathbf{a}_2)$ is nonnegative and either $cov(\mathbf{a}_1, \mathbf{a}_3)$ or $cov(\mathbf{a}_2, \mathbf{a}_3)$ or both are nonpositive for all correlated equilibria.

Now consider games with quadratic utility functions. Quadratic utility functions are often found in applications, for example Cournot oligopoly games with linear demand functions and quadratic costs (see Liu 1996 and Yi 1997 on the uniqueness of correlated equilibria in Cournot oligopoly games and Neyman 1997 on potential games generally). Any game in which best response functions are linear (see for example Manski 1995, p. 116) is naturally represented with quadratic utility functions. Fact 3 says that if a single player has a quadratic utility function, for example $u_1(a_1, a_2) = k_{12}a_1a_2 + k_{11}(a_1)^2 + k_{22}(a_2)^2 + k_1a_1 + k_2a_2 + k_0$, then we can conclude that $k_{12}\, cov_p(\mathbf{a}_1, \mathbf{a}_2) \geq 0$ for all correlated equilibria of the game. This is true regardless of person 2's utility function. So assuming that person 1 has a quadratic utility function and $k_{12} \neq 0$, we have three results. If $k_{12} > 0$, we predict a nonnegative covariance. If we observe positive covariance, we conclude $k_{12} > 0$. If we observe a positive covariance, in all games consistent with this observation, in any correlated equilibrium of any such game, we must have nonnegative covariance.

For another example, say that $n = 3$, $u_1(a_1, a_2, a_3) = a_1a_3 - a_1a_2 - (a_1)^2$, $u_2(a_1, a_2, a_3) = a_1a_2 - a_2a_3 - (a_2)^2$, and $u_3(a_1, a_2, a_3) = -(a_2 - a_3)^2$. By Fact 3, we know that $cov(\mathbf{a}_1, \mathbf{a}_3) - cov(\mathbf{a}_1, \mathbf{a}_2) \geq 0$, $cov(\mathbf{a}_1, \mathbf{a}_2) - cov(\mathbf{a}_2, \mathbf{a}_3) \geq 0$, and $2cov(\mathbf{a}_2, \mathbf{a}_3) \geq 0$. Thus we can conclude that $cov(\mathbf{a}_1, \mathbf{a}_2)$, $cov(\mathbf{a}_1, \mathbf{a}_3)$, and $cov(\mathbf{a}_2, \mathbf{a}_3)$ are all nonnegative in any correlated equilibrium.

Games with incomplete information

In the standard finite game with incomplete information, we have a finite set of people $N = \{1, \ldots, n\}$ and a finite state space Ω. Each person $i \in N$ has a finite strategy set A_i, a prior belief π_i on Ω, a utility function $u_i : A \times \Omega \to \Re$, and a partition $\mathcal{P}_i$ of Ω representing what she knows about the world. Person i's strategy is defined as a function $f_i : \Omega \to \mathcal{A}_i$ which is measurable with respect to $\mathcal{P}_i$, and equilibrium is defined in the standard way. A common example is to say that each person has a "type" in T_i and that each person only knows her own type; in this case $\Omega = \times_{i \in N} T_i$ and $\mathcal{P}_i = \{\{t_i\} \times T_{-i}\}_{t_i \in T_i}$, and a person's strategy can be thought of as a function from T_i to A_i.

To put a game with incomplete information into our framework, for each person $i \in N$ we let $X = A_i$, $Y = A_{-i} \times \Omega$, and $u = u_i$. We thus have a set of constraints IC_i for each person i. The probability distribution p on $A \times \Omega$ is called a mechanism or mediation plan (see for example Myerson 1991), and a mediation plan which satisfies IC_i for all $i \in N$ is called an incentive compatible mediation plan. As is well known, in any equilibrium of the incomplete information game, the resulting distribution over $A \times \Omega$ is an incentive compatible mediation plan. In addition, any equilbrium behavior given any kind of communication device or information sharing results in an incentive compatible mediation plan. Note that $Y = A_{-i} \times \Omega$ and there is no distinction necessary between the actions of other players and exogenous uncertainty.

Here we use a Cournot oligopoly example to contrast our signed covariance approach with the standard equilibrium approach. Say firm 1 produces quantity q_1 and firm 2 produces q_2, and given these quantities, the market price is $60 - (q_1 + q_2)$. Firm 1 has a unit production cost of s_1 and firm 2 has a unit production cost of s_2; these costs randomly vary. Their payoffs are thus given by $u_1(q_1, q_2) = (60 - q_1 - q_2)q_1 - s_1 q_1$ and $u_2(q_1, q_2) = (60 - q_1 - q_2)q_2 - s_2 q_2$.

How do the firm quantities q_1, q_2 depend on each other and on the costs s_1, s_2? The standard way to answer this question is to model the situation as an incomplete information game and find equilibria. Because of the uncertainty about costs s_1, s_2, we must specify the firms' prior beliefs on s_1, s_2 and what they know about s_1, s_2.

The simplest case is when the costs s_1, s_2 are common knowledge, even though they randomly vary. With a bit of calculation, we find equilibrium quantities $q_1^* = 20+s_2/3-2s_1/3$ and $q_2^* = 20 + s_1/3 - 2s_2/3$. This is true regardless of firms' prior beliefs; beliefs are not an issue here since s_1, s_2 are always common knowledge. If we let s_1, s_2 take on values 0, 6, and 12, the equilibrium quantities q_1^*, q_2^* as a function of s_1, s_2 are shown in the table below.

s_1	s_2	q_1^*	q_2^*
0	0	20	20
0	6	22	16
0	12	24	12
6	0	16	22
6	6	18	18
6	12	20	14
12	0	12	24
12	6	14	20
12	12	16	16

How do the results in this table compare with results from our approach? We can write $u_1(q_1, q_2) = 60q_1 - (q_1)^2 - q_2q_1 - s_1q_1$, and thus Fact 3 says that $-cov_p(\mathbf{q}_1, \mathbf{q}_2) - cov_p(\mathbf{q}_1, \mathbf{s}_1) \geq 0$ for any incentive compatible p. Thus $cov_p(\mathbf{q}_1, \mathbf{q}_2) + cov_p(\mathbf{q}_1, \mathbf{s}_1) \leq 0$. In other words, either $cov_p(\mathbf{q}_1, \mathbf{q}_2)$ or $cov_p(\mathbf{q}_1, \mathbf{s}_1)$ or both must be nonpositive. Similarly for firm 2, we get $cov_p(\mathbf{q}_1, \mathbf{q}_2) + cov_p(\mathbf{q}_2, \mathbf{s}_2) \leq 0$.

Note that in the first three rows of the table above, s_1 does not vary. Thus if (s_1, s_2) is distributed only among these three rows, we have $cov_p(\mathbf{q}_1, \mathbf{s}_1) = 0$ and thus we must have $cov_p(\mathbf{q}_1, \mathbf{q}_2) \leq 0$, which is what we observe in the first three rows: as q_1 increases, q_2 decreases. Note that $cov_p(\mathbf{q}_1, \mathbf{q}_2) > 0$ is possible, for example if (s_1, s_2) is distributed only among $(0, 0), (6, 6), (12, 12)$, the first, fifth, and last row of the table. In this case, $cov_p(\mathbf{q}_1, \mathbf{s}_1)$ and $cov_p(\mathbf{q}_2, \mathbf{s}_2)$ must both be negative, which is what we see in the table.

Now consider the case when firm 1 knows only s_1 but firm 2 knows both s_1 and s_2. Now we must specify the firms' prior beliefs: assume that s_1 and s_2 are distributed among $0, 6, 12$, and each of the nine possible states of the world occurs with probability $1/9$. We can compute the Bayesian Nash equilibrium, as shown in the table below. Note that q_1^* does not vary with s_2 because Firm 1 does not know s_2.

s_1	s_2	q_1^*	q_2^*
0	0	22	19
0	6	22	16
0	12	22	13
6	0	18	21
6	6	18	18
6	12	18	15
12	0	14	23
12	6	14	20
12	12	14	17

How do the results in this table compare with our approach? In this equilibrium, and given our prior belief that each state occurs with probability 1/9, we compute $cov_p(\mathbf{q}_1, \mathbf{q}_2) = -16/3$, $cov_p(\mathbf{q}_1, \mathbf{s}_1) = -16$ and $cov_p(\mathbf{q}_2, \mathbf{s}_2) = -12$, consistent with our results that $cov_p(\mathbf{q}_1, \mathbf{q}_2) + cov_p(\mathbf{q}_1, \mathbf{s}_1) \leq 0$ and $cov_p(\mathbf{q}_1, \mathbf{q}_2) + cov_p(\mathbf{q}_2, \mathbf{s}_2) \leq 0$.

Finally, consider the case when firm 1 only knows s_1 and firm 2 only knows s_2. Prior beliefs are as before. The Bayesian Nash equilibrium is shown in the table below.

s_1	s_2	q_1^*	q_2^*
0	0	21	21
0	6	21	18
0	12	21	15
6	0	18	21
6	6	18	18
6	12	18	15
12	0	15	21
12	6	15	18
12	12	15	15

How do the results in this table compare with our approach? In this equilibrium, and given our prior beliefs, we find $cov_p(\mathbf{q}_1, \mathbf{q}_2) = 0$, $cov_p(\mathbf{q}_1, \mathbf{s}_1) = -12$ and $cov_p(\mathbf{q}_2, \mathbf{s}_2) = -12$, again consistent with our results that $cov_p(\mathbf{q}_1, \mathbf{q}_2) + cov_p(\mathbf{q}_1, \mathbf{s}_1) \leq 0$ and $cov_p(\mathbf{q}_1, \mathbf{q}_2) + cov_p(\mathbf{q}_2, \mathbf{s}_2) \leq 0$.

In these three cases, computing equilibria is not difficult, and the equilibria of course provide more precise predictions than the signed covariances. But more complicated scenarios exist. What if the costs s_1 and s_2 are correlated? What if whether a firm knows its own or the other firm's cost depends on what exactly the cost is? What if the firms communicate their costs to each other with noise? What if the firms employ a mediator which recommends actions to take? What if the firms condition their actions on some commonly observed external signal? What if the firms do not have a common prior on s_1, s_2? What if the costs s_1, s_2 are manipulated by some third party with its own objectives? For each possible scenario, we would have to specify explicitly the prior beliefs, the incomplete information, how the mediator sends signals, and so forth, and find one or more equilibria which predict how q_1, q_2 depend on s_1, s_2.

The signed covariance approach makes less precise predictions. But it applies to all the scenarios above, including all possible prior beliefs, all possible specifications of incomplete information, all possible mediation and communication systems, and so forth. By using the IC constraints directly, we can make predictions which hold for all possible scenarios, without any assumptions other than incentive compatibility and the utility function itself.

Another incomplete information game often studied (see Bresnahan and Reiss 1991, Tamer 2003) is below, where u_1, u_2 are considered random "utility shocks."

	0	1
0	$0, 0$	$0, x_2\beta_2 - u_2$
1	$x_1\beta_1 - u_1, 0$	$x_1\beta_1 + \Delta_1 - u_1, x_2\beta_2 + \Delta_2 - u_2$

We have $cov_p(\mathbf{a}_1, (x_1\beta_1 - \mathbf{u}_1)(1 - \mathbf{a}_2) + (x_1\beta_1 + \Delta_1 - \mathbf{u}_1)\mathbf{a}_2) \geq 0$ from the Proposition. But $cov_p(\mathbf{a}_1, (x_1\beta_1 - \mathbf{u}_1)(1 - \mathbf{a}_2) + (x_1\beta_1 + \Delta_1 - \mathbf{u}_1)\mathbf{a}_2) = -cov_p(\mathbf{a}_1, \mathbf{u}_1) + \Delta_1\, cov_p(\mathbf{a}_1, \mathbf{a}_2)$. Thus we have $\Delta_1\, cov_p(\mathbf{a}_1, \mathbf{a}_2) \geq cov_p(\mathbf{a}_1, \mathbf{u}_1)$. Similarly we have $\Delta_2\, cov_p(\mathbf{a}_1, \mathbf{a}_2) \geq cov_p(\mathbf{a}_2, \mathbf{u}_2)$. Thus if $cov_p(\mathbf{a}_1, \mathbf{a}_2) > 0$, we conclude that $\Delta_1 \geq cov_p(\mathbf{a}_1, \mathbf{u}_1)/cov_p(\mathbf{a}_1, \mathbf{a}_2)$ and $\Delta_2 \geq cov_p(\mathbf{a}_2, \mathbf{u}_2)/cov_p(\mathbf{a}_1, \mathbf{a}_2)$.

Thus we can bound Δ_1, Δ_2 simply by computing covariances. In terms of prediction, we can similarly bound $cov_p(\mathbf{a}_1, \mathbf{a}_2)$; for example if Δ_1, Δ_2 have different signs, we have an upper and lower bound for $cov_p(\mathbf{a}_1, \mathbf{a}_2)$. We make these conclusions without any assumption about the distributions of u_1, u_2, the values of $x_1, \beta_1, x_2, \beta_2$, what each person knows about the realizations of u_1, u_2, whether the people can talk to each other, and so forth.

Linear combinations of games

Sometimes it is convenient to express payoffs in a game as the linear combination of payoffs from several simpler games. For example, DeNardo (1995) surveys expert and student preferences over whether the United States and Soviet Union should build weapons systems such as the MX missile, and finds that the great variety of preferences are understandable as convex combinations of certain "strategic extremes" such as the "Pure Dove" and the "Strong Hawk," shown below (the payoffs here are the US's payoffs).

	SU builds	SU doesn't
US builds	1	1
US doesn't	1	4

Pure Dove

	SU builds	SU doesn't
US builds	3	4
US doesn't	1	2

Strong Hawk

Here the Pure Dove prefers for neither side to build the weapon, and any side building the weapon is equally bad. For the Strong Hawk, US superiority is most preferred, both having the weapon is second best, and the worst is for the US to not have the weapon while the Soviet Union does. Let α be the weight given to Pure Dove and $1-\alpha$ be the weight given to Strong Hawk, where $\alpha \in [0,1]$. Say the US is person 1 and the Soviet Union is person 2. Say that building is strategy 1 and not building is strategy 0. In Pure Dove, $u_1(1,\mathbf{a}_2) - u_1(0,\mathbf{a}_2) = -3(1-\mathbf{a}_2)$. In Strong Hawk, $u_1(1,\mathbf{a}_2) - u_1(0,\mathbf{a}_2) = 2$. Hence in the game which is a convex combination of Pure Dove and Strong Hawk, we have $cov_p(\mathbf{a}_1,\mathbf{a}_2) = \alpha cov_p(\mathbf{a}_1, -3(1-\mathbf{a}_2)) + (1-\alpha) cov_p(\mathbf{a}_1, 2) \geq 0$. Thus we get $3\alpha\, cov_p(\mathbf{a}_1,\mathbf{a}_2) \geq 0$. If $\alpha > 0$, we know $cov_p(\mathbf{a}_1,\mathbf{a}_2) \geq 0$. If $\alpha = 0$, in any correlated equilibrium, the US always builds and hence $cov_p(\mathbf{a}_1,\mathbf{a}_2) = 0$. Regardless of what α is, and regardless of the Soviet Union's payoffs, we can conclude that the US and SU actions are nonnegatively correlated.

For another example, say that we observe people playing a 2×2 game. We do not know exactly which game they are playing: possibly chicken, battle of the sexes, matching pennies, or some mixture of the three. Say that the game is a convex combination, with chicken having weight α, battle of sexes having weight β, and matching pennies having weight $1-\alpha-\beta$, as shown below.

	0	1
0	3, 3	1, 4
1	4, 1	0, 0

α

	0	1
0	2, 1	0, 0
1	0, 0	1, 2

β

	0	1
0	1, 0	0, 1
1	0, 1	1, 0

$1-\alpha-\beta$

From the Proposition, we have $\alpha\, cov_p(\mathbf{a}_1, 1-2\mathbf{a}_2)) + \beta\, cov_p(\mathbf{a}_1, -2+3\mathbf{a}_2) + (1-\alpha-\beta) cov_p(\mathbf{a}_1, -1+2\mathbf{a}_2) \geq 0$. Hence $(2-4\alpha+\beta) cov_p(\mathbf{a}_1, \mathbf{a}_2) \geq 0$. From the Proposition we also have $\alpha\, cov_p(\mathbf{a}_2, 1-2\mathbf{a}_1)) + \beta\, cov_p(\mathbf{a}_2, -1+3\mathbf{a}_1) + (1-\alpha-\beta) cov_p(\mathbf{a}_2, 1-2\mathbf{a}_1) \geq 0$. Hence $(5\beta-2) cov_p(\mathbf{a}_1, \mathbf{a}_2) \geq 0$. Thus if we observe a positive covariance, we can conclude that $\beta \geq 2/5$ and $\alpha \leq 3/5$. If we observe a negative covariance, we can conclude that $\beta \leq 2/5$ and $\alpha \geq 3/5$. In other words, a positive covariance indicates that the battle of the sexes "component" is relatively large, while a negative covariance indicates that it is relatively small. This result is intuitive and straightforward, and indeed the question here of estimating α and β given observations should be a simple one. The standard method of finding equilibria requires a much more complicated random utility model.

Local interaction games

A local interaction game can be thought of as each person playing the same 2×2 game with each of his neighbors (see for example Young 1998 and Morris 2000). For each person $i \in N$, let $A_i = \{0,1\}$ and let $N(i) \subset N$ be person i's neighbors (we assume $i \notin N(i)$). Payoffs are defined as $u_i(a) = \sum_{j \in N(i)} v_i(a_i, a_j)$.

The Proposition says that $\sum_{j \in N(i)} cov_p(\mathbf{a}_i, v_i(1, \mathbf{a}_j) - v_i(0, \mathbf{a}_j)) \geq 0$. But we know $v_i(1,\mathbf{a}_j) - v_i(0,\mathbf{a}_j) = (v_i(1,0) - v_i(0,0))(1-\mathbf{a}_j) + (v_i(1,1) - v_i(0,1))\mathbf{a}_j = v_i(1,0) - v_i(0,0) + [v_i(0,0) - v_i(1,0) + v_i(1,1) - v_i(0,1)]\mathbf{a}_j$. Since $v_i(1,0) - v_i(0,0)$ is a constant, we have $[v_i(0,0) - v_i(1,0) + v_i(1,1) - v_i(0,1)] cov_p(\mathbf{a}_i, \sum_{j \in N(i)} \mathbf{a}_j) \geq 0$.

Thus in any local interaction game, given the neighborhood $N(i)$ and the payoffs v_i, we can sign the covariance between a person's action and the sum of his neighbors' actions, as long as $v_i(0,0) - v_i(1,0) + v_i(1,1) - v_i(0,1) \neq 0$. For example, if v_i is a coordination game, with Nash equilibria $(0,0)$ and $(1,1)$, it must be that person i's action is nonnegatively correlated with the sum of her neighbors' actions. Going in the other direction, given observed actions and the neighbors $N(i)$ of person i, we can sign $v_i(0,0) - v_i(1,0) + v_i(1,1) - v_i(0,1)$. Given observed actions and payoffs v_i, we can identify possible sets of neighbors $N(i)$.

On October 1, 3, and 5, 2001, I collected data on whether people in census tract 7016.01 (in Santa Monica, California) displayed flags on their residences, as shown in Figure 1. A plus sign indicates a residence which displays a United States flag or some other red, white, and blue decoration; a dot indicates a residence which does not. There are 1174 total residences in the data set, which is available from the author. The residences in this census tract are primarily single-family homes, although 93 buildings in my data set are multi-unit buildings such as townhouses or apartment buildings. A data point here is an individual building; for example, when a flag appears on an apartment building, the entire building is counted as displaying a flag and no attempt is made to figure out which apartment in the building is displaying the flag and which ones are not. Only residential buildings are included. According to the 2000 US Census, 3957 people live in this census tract and there are a total of 1863 housing units.

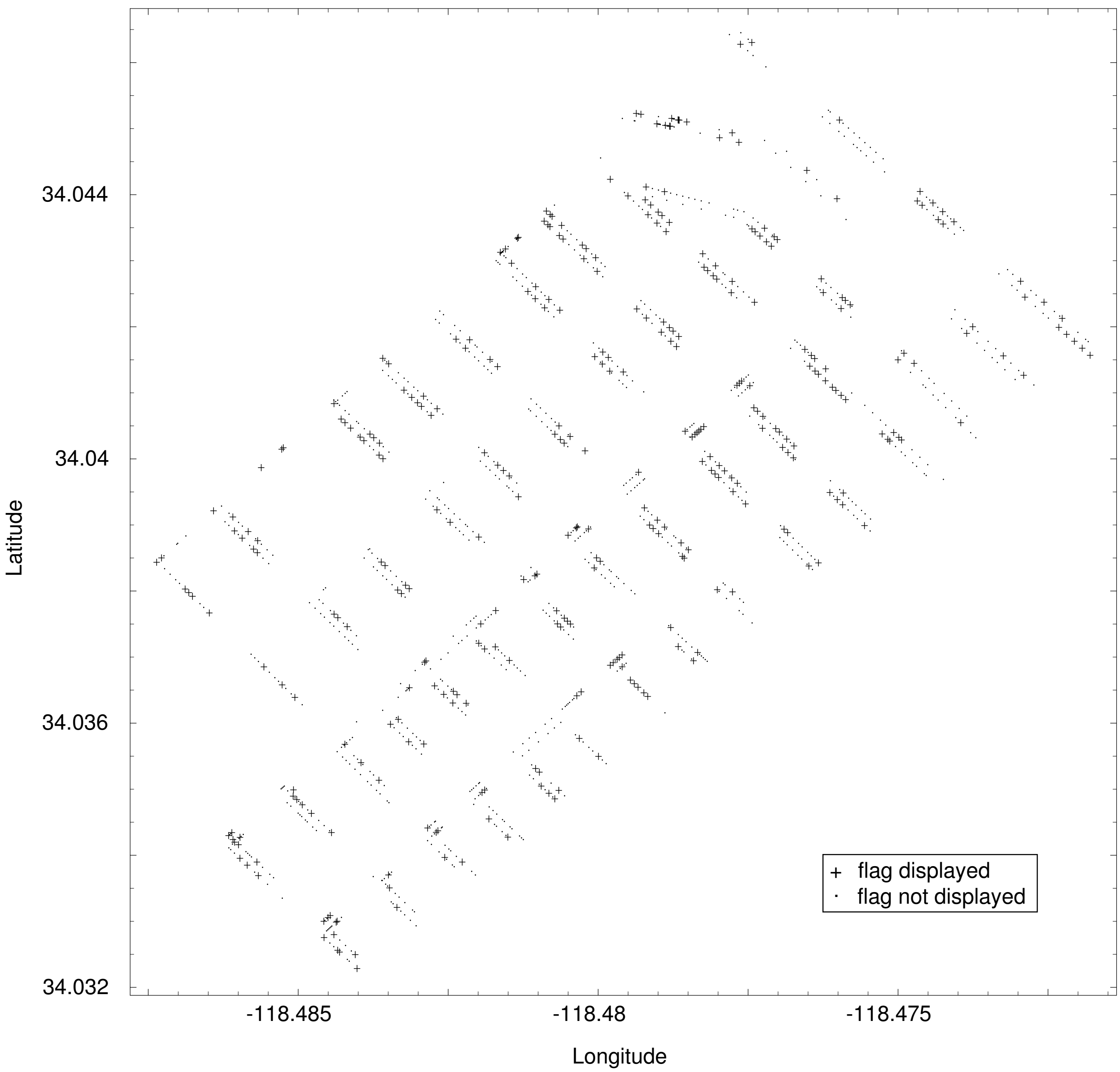

Figure 1. Flag display in census tract 7016.01
(Santa Monica, California), October 1, 3, 5, 2001

In my data, 362 of the 1174 residences (30.8 percent) display flags. Inspecting Figure 1, it seems that a person's choice of whether to display a flag depends on whether her neighbors display a flag; for example, there are some blocks in which nearly everyone displays a flag, which would be unlikely if people's decisions were independent.

We can model this as a local interaction game, where putting up a flag is strategy 1 and not putting up one is strategy 0, and payoffs are $v(0,0), v(0,1), v(1,0), v(1,1)$ (assume these payoffs are the same for everyone). Let $N(i)$, the neighbors of i, be the houses on the same block adjacent to i. In our data, 947 of the 1174 residences have two neighbors in this sense, 220 have one neighbor, and 7 have no neighbors. We find that $cov_p(\mathbf{a}_i, \sum_{j \in N(i)} \mathbf{a}_j) = (250/1174) - (362/1174)(655/1174) \approx 0.0409$; in other words, the covariance between a person's action and the actions of his neighbors is positive. Hence $v(0,0) - v(1,0) + v(1,1) - v(0,1) \geq 0$. Since both strategy 1 and strategy 0 are observed, we assume that neither is strongly dominated, and hence we conclude that $v(0,0) \geq v(1,0)$ and $v(1,1) \geq v(0,1)$, or in other words, v is a coordination game with two pure Nash equilibria $(0,0)$ and $(1,1)$.

To identify $v(0,0), v(0,1), v(1,0), v(1,1)$ more precisely, we can directly use the IC inequalities. Of the residences which put up flags, on average $250/362 \approx 0.691$ of their neighbors also put up flags and $405/362 \approx 1.119$ of their neighbors do not put up flags. Hence we have the inequality $(250/362)v(1,1) + (405/362)v(1,0) \geq (250/362)v(0,1) + (405/362)v(0,0)$. Of the residences which do not put up flags, on average $405/812 \approx 0.499$ of their neighbors put up flags and $1054/812 \approx 1.298$ of their neighbors do not put up flags. Hence we have the inequality $(405/812)v(0,1) + (1054/812)v(0,0) \geq (405/812)v(1,1) + (1054/812)v(1,0)$. We can normalize $v(0,1) = v(1,0) = 0$; assuming that $v(0,0) \neq 0$, we can also normalize $v(0,0) = 1$. We thus have $v(1,1) \in [405/250, 1054/450] \approx [1.620, 2.342]$. Note that if we assume Nash equilibrium instead of the IC inequalities, we cannot identify the magnitude of $v(1,1)$ because any positive $v(1,1)$ is consistent with $(1,1)$ being a Nash equilibrium.

We might suppose a more complicated model; for example a person's immediate neighbors might affect her payoff more than people who live two houses away. Say that person i has immediate neighbors $N(i)$ and peripheral neighbors $NN(i)$. Say that a person gets payoffs $v(0,0), v(0,1), v(1,0), v(1,1)$ from immediate neighbors and payoffs $vv(0,0), vv(0,1), vv(1,0), vv(1,1)$ from peripheral neighbors, and assume for convenience that $v(0,1) = v(1,0) = vv(0,1) = vv(1,0) = 0$. So if a person puts up a flag and one of her immediate neighbors and two of her peripheral neighbors put up flags, she gets payoff $v(1,1) + 2vv(1,1)$ for example. From the Proposition, we know that $(v(0,0) + v(1,1))cov_p(\mathbf{a}_i, \sum_{j \in N(i)} \mathbf{a}_j) + (vv(0,0) + vv(1,1))cov_p(\mathbf{a}_i, \sum_{j \in NN(i)} \mathbf{a}_j) \geq 0$. Let $N(i)$ again be the houses on the same block adjacent to i and let $NN(i)$ be the houses on the same block two houses away from i on either side. In the data, we have $cov_p(\mathbf{a}_i, \sum_{j \in N(i)} \mathbf{a}_j) \approx 0.0409$ and $cov_p(\mathbf{a}_i, \sum_{j \in NN(i)} \mathbf{a}_j) \approx 0.0278$. Thus $0.0409(v(0,0)+v(1,1))+0.0278(vv(0,0)+vv(1,1)) \geq 0$. This it cannot be that both v and vv are "anticoordination" games (games in which $(0,1)$ and $(1,0)$ are the pure Nash equilibria). If we assume $v(0,0) = v(1,1)$ and $vv(0,0) = vv(1,1)$ for simplicity, we have $0.0409v(0,0) + 0.0278vv(0,0) \geq 0$. If vv is an anticoordination game with $vv(0,0) = -1$, then v must be at least a "weak" coordination game with $v(0,0) \geq 0.680$. If v is an anticoordination game with $v(0,0) = -1$, then vv must be a relatively "strong" coordination game with $vv(0,0) \geq 1.471$.

Conclusion: statistical game theory

The most common solution concepts in game theory and choice theory are "point" predictions which predict a single action for each person, as in Nash equilibrium. However, when we analyze data, we typically look not at each particular data point by itself but at the statistical relationships among different variables given all the data. This paper works toward a "statistical game theory" which makes predictions given a game, and identifies games given data, fundamentally in terms of statistical relationships. This paper demonstrates a relationship between the fundamental game-theoretic concept of incentive compatibility (revealed preference in choice-theoretic terms) and the fundamental statistical concept of covariance.

Most work on applying games to empirical data is based on point predictions such as Nash equilibrium. Thus explaining observed variation requires adding exogenous randomness or heterogeneity, for example by allowing random mistakes or by allowing payoffs in a game to vary randomly (for example Bresnahan and Reiss 1991, Lewis and Schultz 2003, McKelvey and Palfrey 1995, Signorino 2003, Tamer 2003). This paper shares the motivations of this literature, and as explored in earlier examples, has direct application for example to models with random utilities. This paper, however, derives statistical relationships not from randomness added on to a game but from what is inherent in the game itself, assuming nothing more than incentive compatibility.

There are three advantages to using incentive compatibility constraints instead of a point-prediction approach. First, since the set of incentive compatible distributions is convex, any aggregation of incentive compatible distributions is also incentive compatible. For example, if each group of individuals in a population is playing an incentive compatible distribution, or if each play in a sequence of trials is an incentive compatible distribution, then the distribution aggregated over the population or over the sequence is incentive compatible. One need not worry about whether a given observation is an individual or aggregate, since they can be treated in the same way. This is not usually true for point predictions; for example, the set of Nash equilibria is not convex.

Second, the assumptions here are much weaker than the assumptions behind Nash equilibrium. All Nash equilibria of any game, including games of incomplete information, satisfy the incentive compatibility constraints, and all of the results in this paper still hold if we more traditionally assume Nash equilibria and their mixtures. If people play a mixture of several pure strategy and mixed strategy Nash equilibria, instead of figuring out how this mixture can result from the various Nash equilibria, we can simply use the mixture directly and calculate covariances to identify the game.

Third, typically the existence of multiple equilibria is considered a shortcoming which needs to be somehow fixed. For example, when applying games to empirical data, because of multiple equilibria, the distribution of equilibria cannot be exactly determined by assumptions about how the game is distributed (see for example Tamer 2003). There has been much work in game theory developing criteria for selecting one of several equilibria. The statistical approach here completely avoids this issue, and considers not one but all possible incentive compatible distributions. The statistical approach is concerned not with any single predicted action but with statistical relationships among actions; a wide range of possible actions is something to be embraced, not avoided.

Incentive compatibility is a fundamental concept, and since it is defined as a set of linear inequalities, it is mathematically simple compared to Nash equilibrium for example, which is the fixed point of a correspondence. However, even people familiar with the concept find it difficult to intuitively “visualize,” even in the simplest case of a 2×2 game (see Nau, Gomez-Canovas, and Hansen 2004 and Calvó-Armengol 2003). This paper shows how incentive compatibility can be understood in “reduced form” as a signed covariance.

References

Aumann, Robert J. 1974. "Subjectivity and Correlation in Randomized Strategies." *Journal of Mathematical Economics* 1: 67–96.

Bresnahan, Timothy F. and Peter C. Reiss. 1991. "Empirical Models of Discrete Games." *Journal of Econometrics* 48: 57–81.

Calvó-Armengol, Antoni. 2003. "The Set of Correlated Equilibria of 2×2 Games." Working paper, Universitat Autònoma de Barcelona.

DeNardo, James. 1995. *The Amateur Strategist: Intuitive Deterrence Theories and the Politics of the Nuclear Arms Race.* Cambridge: Cambridge University Press.

Fortuin, C. M., Kasteleyn, P. W., and J. Ginibre. 1971. "Correlation Inequalities on Some Partially Ordered Sets." *Communications in Mathematical Physics* 22: 89–103.

Hart, Sergiu and David Schmeidler. 1989. "Existence of Correlated Equilibria." *Mathematics of Operations Research* 14: 18–25.

Karlin, Samuel and Yosef Rinott. 1980a. "Classes of Orderings of Measures and Related Correlation Inequalities. I. Multivariate Totally Positive Distributions." *Journal of Multivariate Analysis* 10: 467–498.

Karlin, Samuel and Yosef Rinott. 1980b. "Classes of Orderings of Measures and Related Correlation Inequalities. II. Multivariate Reverse Rule Distributions." *Journal of Multivariate Analysis* 10: 499–516.

Lewis, Jeffrey B. and Kenneth A. Schultz. 2003. "Revealing Preferences: Empirical Estimation of a Crisis Bargaining Game with Incomplete Information." *Political Analysis* 11: 345–367.

Liu, Luchuan. 1996. "Correlated Equilibrium of Cournot Oligopoly Competition." *Journal of Economic Theory* 68: 544–548.

Manski, Charles F. 1995. *Identification Problems in the Social Sciences.* Cambridge, Massachusetts: Harvard University Press.

McKelvey, Richard D. and Thomas R. Palfrey. 1995. "Quantal Response Equilibria for Normal Form Games." *Games and Economic Behavior* 10: 6–38.

Milgrom, Paul R. and Robert J. Weber. 1982. "A Theory of Auctions and Competitive Bidding." *Econometrica* 50: 1089–1122.

Moreno, Diego and John Wooders. 1998. "An Experimental Study of Communication and Coordination in Noncooperative Games." *Games and Economic Behavior* 24: 47–76.

Morris, Stephen. 2000. "Contagion." *Review of Economic Studies* 67: 57–78.

Myerson, Roger B. 1991. *Game Theory: Analysis of Conflict.* Cambridge, Massachusetts: Harvard University Press.

Nau, Robert, Sabrina Gomez-Canovas, and Pierre Hansen. 2004. "On the Geometry of Nash Equilibria and Correlated Equilibria." *International Journal of Game Theory* 32: 443–453.

Nau, Robert F. and Kevin F. McCardle. 1990. "Coherent Behavior in Noncooperative Games." *Journal of Economic Theory* 50: 424–444.

Neyman, Abraham. 1997. "Correlated Equilibrium and Potential Games." *International Journal of Game Theory* 26: 223–227.

Signorino, Curtis S. 2003. "Structure and Uncertainty in Discrete Choice Models." *Political Analysis* 11: 316–344.

Tamer, Elie. 2003. "Incomplete Simultaneous Discrete Response Model with Multiple Equilibria." *Review of Economic Studies* 70: 147–165.

Yi, Sang-Seung. 1997. "On the Existence of a Unique Correlated Equilibrium in Cournot Oligopoly," *Economics Letters* 54: 237–241.

Young, H. Peyton. 1998. *Individual Strategy and Social Structure: An Evolutionary Theory of Institutions.* Princeton: Princeton University Press.

Appendix

To prove the Proposition, we first define $\phi((x,y),(x',y'))$ and derive two lemmas. Given $p \in P$ and $(x,y),(x',y') \in X \times Y$, define $\phi((x,y),(x',y')) = p(x,y)p(x',y') - p(x,y')p(x',y)$. Lemma 1 says that the incentive compatibility constraints imply a linear inequality on $\phi((x,y),(x',y'))$.

Lemma 1. If p, u satisfy IC, then $\sum_{y \in Y}(u(x,y) - u(x',y))\phi((x,y),(x',y')) \geq 0$.

Proof. We write IC as $\sum_{y \in Y} p(x,y)(u(x,y) - u(x',y)) \geq 0$. Multiplying both sides by $p(x',y')$, we have $\sum_{y \in Y} p(x,y)p(x',y')(u(x,y) - u(x',y)) \geq 0$. We call this inequality $(*)$. Similarly, we have the IC inequality $\sum_{y \in Y} p(x',y)(u(x',y) - u(x,y)) \geq 0$. Multiplying both sides by $p(x,y')$, we have $\sum_{y \in Y} p(x,y')p(x',y)(u(x',y) - u(x,y)) \geq 0$. We call this inequality $(**)$. Add the inequalities $(*)$ and $(**)$ together and we are done. ■

Lemma 2 says that the covariance of two random variables is a linear function of the $\phi((x,y),(x',y'))$. Lemma 2 is well known (see for example C. M. Fortuin, P. W. Kasteleyn, and J. Ginibre 1971), but we state and prove it here for the sake of completeness. We say a random variable $\mathbf{f} : X \times Y \to \Re$ is constant in y if $\mathbf{f}(x,y) = \mathbf{f}(x,y')$ for all $x \in X$, $y, y' \in Y$. We say a random variable $\mathbf{g} : X \times Y \to \Re$ is constant in x if $\mathbf{g}(x,y) = \mathbf{g}(x',y)$ for all $x, x' \in X$, $y \in Y$.

Lemma 2. Say that $\mathbf{f} : X \times Y \to \Re$ is constant in y and $\mathbf{g} : X \times Y \to \Re$ is constant in x. Say that $Z = X^0 \times Y^0$, where $X^0 \subset X$, $Y^0 \subset Y$, and $p(Z) > 0$. Then $cov_p(\mathbf{f}, \mathbf{g}|Z) = [1/(4p(Z)^2)] \sum_{(x,y),(x',y') \in Z}(\mathbf{f}(x,y) - \mathbf{f}(x',y'))(\mathbf{g}(x,y) - \mathbf{g}(x',y'))\phi((x,y),(x',y'))$.

Proof. Let $k = \sum_{(x,y),(x',y') \in Z}(\mathbf{f}(x,y) - \mathbf{f}(x',y'))(\mathbf{g}(x,y) - \mathbf{g}(x',y'))p(x,y)p(x',y')$. We expand this and get $k = \sum \mathbf{f}(x,y)\mathbf{g}(x,y)p(x,y)p(x',y') - \sum \mathbf{f}(x,y)\mathbf{g}(x',y')p(x,y)p(x',y') - \sum \mathbf{f}(x',y')\mathbf{g}(x,y)p(x,y)p(x',y') + \sum \mathbf{f}(x',y')\mathbf{g}(x',y')p(x,y)p(x,'y') =$
$p(Z)^2[E_p(\mathbf{fg}|Z) - E_p(\mathbf{f}|Z)E_p(\mathbf{g}|Z) - E_p(\mathbf{f}|Z)E_p(\mathbf{g}|Z) + E_p(\mathbf{fg}|Z)] = 2p(Z)^2 cov_p(\mathbf{f},\mathbf{g}|Z)$. Let there be a one-to-one function $m : X^0 \to \{1, 2, \ldots, \#X^0\}$ which assigns a unique number to each member of X^0. Note that $k = (\sum_{(x,y),(x',y') \in Z : m(x) > m(x')} + \sum_{(x,y),(x',y') \in Z : m(x) < m(x')}$
$+ \sum_{(x,y),(x',y') \in Z : m(x) = m(x')})(\mathbf{f}(x,y) - \mathbf{f}(x',y'))(\mathbf{g}(x,y) - \mathbf{g}(x',y'))p(x,y)p(x',y')$. The third sum is zero because $\mathbf{f}$ is constant in y and $m(x) = m(x')$ implies $x = x'$.

If we let $(v,w)=(x',y)$ and $(v',w')=(x,y')$, then the second sum can be written as $\sum_{(v,w),(v',w')\in Z:m(v)>m(v')} -(\mathbf{f}(v,w)-\mathbf{f}(v',w'))(\mathbf{g}(v,w)-\mathbf{g}(v',w'))p(v',w)p(v,w')$, because $\mathbf{f}$ is constant in y and $\mathbf{g}$ is constant in x. Changing variables again, where $(x,y)=(v,w)$ and $(x',y')=(v',w')$, the second sum is $\sum_{(x,y),(x',y')\in Z:m(x)>m(x')}(\mathbf{f}(x,y)-\mathbf{f}(x',y'))(\mathbf{g}(x,y)-\mathbf{g}(x',y'))(-p(x',y)p(x,y'))$. Thus $k=\sum_{(x,y),(x',y')\in Z:m(x)>m(x')}(\mathbf{f}(x,y)-\mathbf{f}(x',y'))(\mathbf{g}(x,y)-\mathbf{g}(x',y'))\phi((x,y),(x',y')) = (1/2)\sum_{(x,y),(x',y')\in Z}(\mathbf{f}(x,y)-\mathbf{f}(x',y'))(\mathbf{g}(x,y)-\mathbf{g}(x',y'))$ $\phi((x,y),(x',y'))$. Since $k=2p(Z)^2 cov_p(\mathbf{f},\mathbf{g}|Z)$, we are done. ■

Proposition. Say p,u satisfy IC and $x,x'\in X$. Then

$$cov_p(\mathbf{1}_x, u(x,\mathbf{y})-u(x',\mathbf{y})|\{x,x'\}\times Y)\geq 0.$$

Proof. If $p(\{x,x'\}\times Y)=0$, we are done, by our convention that $cov_p(\mathbf{f},\mathbf{g}|Z)=0$ if $p(Z)=0$. So assume that $p(\{x,x'\}\times Y)>0$. Lemma 2 says $cov_p(\mathbf{1}_x,u(x,\mathbf{y})-u(x',\mathbf{y})|\{x,x'\}\times Y)=[1/(4p(\{x,x'\}\times Y)^2](1-0)\sum_{y,y'\in Y}[(u(x,y)-u(x',y))-(u(x,y')-u(x',y'))]\phi((x,y),(x',y'))$. It suffices to show that this sum is nonnegative.

From Lemma 1, we know that $\sum_{y\in Y}(u(x,y)-u(x',y))\phi((x,y),(x',y'))\geq 0$. Hence $\sum_{y,y'\in Y}(u(x,y)-u(x',y))\phi((x,y),(x',y'))\geq 0$. Call this inequality $(*)$. From the definition of $\phi((x,y),(x',y'))$, we have $\phi((x,y),(x',y'))=-\phi((x',y),(x,y'))$. Hence $\sum_{y,y'\in Y}-(u(x,y)-u(x',y))\phi((x',y),(x,y'))\geq 0$. If we change variables and let $y'=y$ and $y=y'$, we have $\sum_{y,y'\in Y}-(u(x,y')-u(x',y'))\phi((x',y'),(x,y))\geq 0$. Since $\phi((x',y'),(x,y))=\phi((x,y),(x',y'))$, we have $\sum_{y,y'\in Y}-(u(x,y')-u(x',y'))\phi((x,y),(x',y'))\geq 0$. Call this inequality $(**)$. Add $(*)$ and $(**)$ together and we are done. ■

Fact 1. Say $X=\{x,x'\}$ and $Y=\{y,y'\}$ and u is nontrivial. Then either $cov_p(\mathbf{1}_x,\mathbf{1}_y)\geq 0$ for all $p\in IC(u)$ or $cov_p(\mathbf{1}_x,\mathbf{1}_y)\leq 0$ for all $p\in IC(u)$.

Proof. By the Proposition, we have $cov_p(\mathbf{1}_x,u(x,\mathbf{y})-u(x',\mathbf{y}))\geq 0$. It is easy to verify that $u(x,\mathbf{y})-u(x',\mathbf{y})=(u(x,y)-u(x',y))\mathbf{1}_y+(u(x,y')-u(x',y'))(1-\mathbf{1}_y)$. Hence $cov_p(\mathbf{1}_x,(u(x,y)-u(x',y))\mathbf{1}_y+(u(x,y')-u(x',y'))(1-\mathbf{1}_y))=(u(x,y)-u(x',y)-u(x,y')+u(x',y'))cov_p(\mathbf{1}_x,\mathbf{1}_y)\geq 0$. If $u(x,y)-u(x',y)-u(x,y')+u(x',y')\neq 0$, we are done. Assume $u(x,y)-u(x',y)-u(x,y')+u(x',y')=0$. If $u(x,y)>u(x',y)$, then $u(x,y')>u(x',y')$

and thus IC implies $p(x', y) = p(x', y') = 0$, and thus $cov_p(\mathbf{1}_x, \mathbf{1}_y) = 0$. Similarly, if $u(x, y) < u(x', y)$, we have $cov_p(\mathbf{1}_x, \mathbf{1}_y) = 0$. Thus we are left with the case when $u(x, y) = u(x', y)$, and thus $u(x, y') = u(x', y')$, in which case u is trivial. ■

To prove Fact 3, we need Lemma 3.

Lemma 3. Say that $X \subset \Re$ and that $\mathbf{f} : X \times Y \to \Re$ is constant in y and $\mathbf{g} : X \times Y \to \Re$ is constant in x. Then $cov_p(\mathbf{f}, \mathbf{g}) = \sum_{x,x' \in X, x > x'} p(\{x, x'\} \times Y)^2 cov_p(\mathbf{f}, \mathbf{g} | \{x, x'\} \times Y)$.

Proof. By Lemma 2, we have $cov_p(\mathbf{f}, \mathbf{g}) = (1/4) \sum_{(x,y),(x',y')} (\mathbf{f}(x, y) - \mathbf{f}(x', y'))(\mathbf{g}(x, y) - \mathbf{g}(x', y'))\phi((x, y), (x', y')) = (1/4)(\sum_{(x,y),(x',y'),x>x'} + \sum_{(x,y),(x',y'),x=x'} + \sum_{(x,y),(x',y'),x<x'})$. Since $\mathbf{f}$ is constant in y, the second sum is zero. Show that the first sum and the third sum are equal. If we change variables ($x = x'$ and $x' = x$), the third sum is $\sum_{x'<x} (\mathbf{f}(x', y) - \mathbf{f}(x, y'))(\mathbf{g}(x', y) - \mathbf{g}(x, y'))\ \phi((x', y), (x, y'))$. Since $\phi((x', y), (x, y')) = -\phi((x, y), (x', y'))$, the third sum is $\sum_{x'<x} (\mathbf{f}(x, y') - \mathbf{f}(x', y))(\mathbf{g}(x', y) - \mathbf{g}(x, y'))\phi((x, y), (x', y'))$, which is equal to the first sum since $\mathbf{f}$ is constant in y and $\mathbf{g}$ is constant in x. Hence $cov_p(\mathbf{f}, \mathbf{g}) = (1/2) \sum_{x>x'} (\mathbf{f}(x, y) - \mathbf{f}(x', y'))(\mathbf{g}(x, y) - \mathbf{g}(x', y'))\phi((x, y), (x', y'))$.

By Lemma 2, $cov_p(\mathbf{f}, \mathbf{g} | \{x, x'\} \times Y) = [1/(4p(\{x, x'\} \times Y)^2)] \sum_{y,y' \in Y} [(\mathbf{f}(x, y) - \mathbf{f}(x', y'))(\mathbf{g}(x, y) - \mathbf{g}(x', y'))\phi((x, y), (x', y')) + (\mathbf{f}(x', y) - \mathbf{f}(x, y'))(\mathbf{g}(x', y) - \mathbf{g}(x, y'))\phi((x', y), (x, y'))] = [1/(2p(\{x, x'\} \times Y)^2)] \sum_{y,y' \in Y} (\mathbf{f}(x, y) - \mathbf{f}(x', y'))(\mathbf{g}(x, y) - \mathbf{g}(x', y'))\phi((x, y), (x', y'))$. Therefore $cov_p(\mathbf{f}, \mathbf{g}) = \sum_{x>x'} p(\{x, x'\} \times Y)^2 cov_p(\mathbf{f}, \mathbf{g} | \{x, x'\} \times Y)$. ■

Fact 3. Say that $X \subset \Re$, $Y \subset \Re^m$ and u satisfies the condition that $u(x, y) - u(x', y) = v(x, x') \sum_{j=1}^m c_j y_j + w(x, x')$, where $c_j \in \Re$ and $v(x, x') > 0$ when $x > x'$. Say p, u satisfy IC. Then $\sum_{j=1}^m c_j\, cov_p(\mathbf{x}, \mathbf{y}_j) \geq 0$.

Proof. By the Proposition, we know that $cov_p(\mathbf{1}_x, v(x, x') \sum_{j=1}^m c_j \mathbf{y}_j + w(x, x') | \{x, x'\} \times Y) \geq 0$. Since $v(x, x') > 0$ and $w(x, x')$ are constants, we have $cov_p(\mathbf{1}_x, \sum_{j=1}^m c_j \mathbf{y}_j | \{x, x'\} \times Y) \geq 0$. Hence when $x > x'$, we have $cov_p(\mathbf{x}, \sum_{j=1}^m c_j \mathbf{y}_j | \{x, x'\} \times Y) \geq 0$. By Lemma 3, we have $cov_p(\mathbf{x}, \sum_{j=1}^m c_j \mathbf{y}_j) = \sum_{x>x'} p(\{x, x'\} \times Y)^2 cov_p(\mathbf{x}, \sum_{j=1}^m c_j \mathbf{y}_j | \{x, x'\} \times Y) \geq 0$. Hence $\sum_{j=1}^m c_j\, cov_p(\mathbf{x}, \mathbf{y}_j) = cov_p(\mathbf{x}, \sum_{j=1}^m c_j \mathbf{y}_j) \geq 0$. ■

Fact 4. Say that p, u satisfy IC and $x, x' \in X$. Say that $u(x,y) - u(x',y)$ is not constant in y and that $p(x',y) > 0$ for $y \in Y$. The following statement is not true:

$$u(x,y) - u(x',y) > u(x,y') - u(x',y') \Leftrightarrow p(x,y)/p(x',y) < p(x,y')/p(x',y').$$

Proof. We know $\sum_{y,y' \in Y}[(u(x,y) - u(x',y)) - (u(x,y') - u(x',y'))]\phi((x,y),(x',y')) \geq 0$ from the proof of the Proposition. If the statement in Fact 4 is true, then $(u(x,y) - u(x',y)) - (u(x,y') - u(x',y')) > 0 \Leftrightarrow p(x,y)/p(x',y) < p(x,y')/p(x',y') \Leftrightarrow \phi((x,y),(x',y')) < 0$. Since $u(x,y) - u(x',y)$ is not constant in y, it must be that $\sum_{y,y' \in Y}[(u(x,y) - u(x',y)) - (u(x,y') - u(x',y'))]\phi((x,y),(x',y')) < 0$, a contradiction. ■

Fact 5. Say $n = 2$, $A_1 = \{a_1, b_1\}$, $A_2 = \{a_2, b_2\}$, and $p \in CE(u)$. If $cov_p(\mathbf{1}_{a_1}, \mathbf{1}_{a_2}) > 0$, then (a_1, a_2) and (b_1, b_2) are Nash equilibria of u. If $cov_p(\mathbf{1}_{a_1}, \mathbf{1}_{a_2}) < 0$, then (a_1, b_2) and (b_1, a_2) are Nash equilibria of u.

Proof. By the Proposition, we have $cov_p(\mathbf{1}_{a_1}, u_1(a_1, \mathbf{a}_2) - u_1(b_1, \mathbf{a}_2)) \geq 0$. It is easy to verify that $u_1(a_1, \mathbf{a}_2) - u_1(b_1, \mathbf{a}_2) = (u_1(a_1, a_2) - u_1(b_1, a_2))\mathbf{1}_{a_2} + (u_1(a_1, b_2) - u_1(b_1, b_2))(1 - \mathbf{1}_{a_2})$, and thus $cov_p(\mathbf{1}_{a_1}, (u_1(a_1, a_2) - u_1(b_1, a_2))\mathbf{1}_{a_2} + (u_1(a_1, b_2) - u_1(b_1, b_2))(1 - \mathbf{1}_{a_2})) = (u_1(a_1, a_2) - u_1(b_1, a_2) - u_1(a_1, b_2) + u_1(b_1, b_2))cov_p(\mathbf{1}_{a_1}, \mathbf{1}_{a_2}) \geq 0$. Thus if $cov_p(\mathbf{1}_{a_1}, \mathbf{1}_{a_2}) > 0$, we have $u_1(a_1, a_2) - u_1(b_1, a_2) - u_1(a_1, b_2) + u_1(b_1, b_2) \geq 0$. It cannot be that person 1 has a strongly dominated strategy (because then $cov_p(\mathbf{1}_{a_1}, \mathbf{1}_{a_2}) = 0$) and thus $u_1(a_1, a_2) \geq u_1(b_1, a_2)$ and $u_1(b_1, b_2) \geq u_1(a_1, b_2)$. We show $u_2(a_1, a_2) \geq u_2(a_1, b_2)$ and $u_2(b_1, b_2) \geq u_2(b_1, a_2)$ similarly. If $cov_p(\mathbf{1}_{a_1}, \mathbf{1}_{a_2}) < 0$, the argument is similar. ■