CAUSAL ASSESSMENT, EQUILIBRIUM AND AMBIGUITY IN GAMES

By Michael D. Ryall*
Carol Simon Hall
University of Rochester,
Rochester, NY 14627
ryall@ssb.rochester.edu

December 9, 1999

This paper presents a treatment of games in which players utilize and assess causal theories regarding the nature of interactions in their environment. I define a *causal form game,* in which potential and actual causal relationships have explicit representation. A key technical result is imported from artificial intelligence theory linking a game's causal representation to the stochastic behavior arising from the behavior of its players.

In a *causal equilibrium*, players' beliefs regarding causal theories are consistent with the data generated during play. A specific measure of "causal ambiguity" is proposed and demonstrated to be only loosely related to commonly cited measures of environmental complexity.

Keywords: Causality, ambiguity, causal schema, subjective equilibrium.

*I thank Bryan Ellickson, David Levine, Colin Camerer, David Chickering, Steven Lippman, Glen MacDonald, Leslie Marx, John Miller, Joseph Ostroy, Scott Page, Juda Pearl, Jose Penalva, Eduardo Zambrano and Bill Zame for thoughtful comments and the John M. Olin foundation for financial support.

1. Introduction

The search for causal explanations of real world phenomena appears to be a salient characteristic of human nature. As a result, it is not surprising that the study of causal inference spans many academic disciplines. Philosophers have long been interested in the nature of causality (at least since the appearance of David Hume's *A Treatise of Human Nature* in 1739). More recently, behavioral decision theorists, such as Tversky and Kahneman (1977) and Einhorn and Hogarth (1986), argue that when individuals face important decisions in a complex and uncertain environment, they typically begin by constructing a mental mapping of probable causes and likely consequences (such maps are referred to as "causal schema"). Computer scientists Pearl (1988), Verma and Pearl (1992), and Chickering (1996) use causal models as a basis for artificial intelligence. Statisticians use causal ideas to test how useful some variables are for forecasting others (see Granger, 1969 and Sims, 1972). In economics, Lippman and Rumelt (1982) demonstrate that persistent interfirm performance differences can arise when causal ambiguity with respect to a firm's production process (i.e., when its workings remain hidden to outsiders) impedes imitation.

The behavioral literature does not appear to address the issue of consistency between one's causal model and the observed behavior of the environment, presumably an important concern in the context of learning. And, while Lippman and Rumelt appeal to the notion of causal ambiguity to justify their assumption of impeded imitability, their model does not specify causal ambiguity per se. In this paper, Kalai and Lehrer's more general subjective environmental response functions are replaced with *causal theories* that provide agents both with a basis for subjective optimization as well as with a forecast of expected outcomes. In so doing, it refines subjective equilibrium with the notion of a *causal equilibrium*, provides an explicit interpretation of causal inference in games, defines confirmability consistency with respect to such inferences and introduces a specific measure of causal ambiguity.

One issue central to much of the current research in non-cooperative game theory is the identification of equilibria that result from long-run learning processes. When equilibrium is interpreted as the outcome of less than fully rational individuals struggling to optimize given their experiences – as opposed to the result of rational pre-play analysis and introspection – it is often necessary to expand the set of candidate outcomes beyond those implied by Nash. Several authors have proposed equilibrium ideas along these lines including, for example, Pearce's (1984) rationalizability, Battigalli and Guatoli's (1988) conjectural equilibrium, Rubinstein and Wolinsky's (1990) rationalizable conjectural equilibrium, Fudenberg and Levine's (1993) self-confirming equilibrium and Kalai and Lehrer's (1993, 1995) subjective equilibrium. These concepts weaken Nash by allowing various types of errors to find their way into player assessments.[1]

For example, Fudenberg and Levine (1998) say, "The most natural assumption in many such contexts is that agents observe the terminal nodes

[1]For related discussions, see Abreu, Pearce, and Stacchetti (1990), Blume and Easley (1992), Brandenberger (1993), Geanakopolis (1994), Harsanyi (1967-68), Milgrom and Roberts (1990), and Nachbar (1996). Howitt and McAfee (1992) employ a similar idea in a macroeconomic application.

(outcomes) that are reached in their own plays of the game, but that agents do not observe the parts of their opponent's strategies that specify how opponents would have played at information sets that are not reached in that play of the game."

Nash equilibrium requires players with accurate knowledge about the structure of the game to implement strategies that are best responses to exactly correct beliefs about the play of their opponents. Most coarsening of Nash maintain the assumptions of structural knowledge and optimizing behavior while relaxing the requirement that beliefs regarding opponent play be perfect. These notions typically impose various degrees of belief consistency along two key dimensions: rationalizability and confirmability. A player's beliefs regarding opponents' play are *rationalizable* when they are consistent with his knowledge both about the structure of the game as well as the rationality of his opponents. His beliefs are *confirmable* when they are consistent with the objective distribution of outcomes along the equilibrium path of play.

Since the number of rational strategies available to a player is generally greater than one, rationalizable beliefs can be wrong when they place positive probability on rational opponent actions that are not actually played. Confirmable beliefs can be wrong regarding off-equilibrium path play; i.e., regarding counterfactual events. In some cases, rationalizable and confirmable constraints are used in conjunction with one another (as is the case with rationalizable conjectural equilibrium).[2]

[2]In a personal communication, G. MacDonald points out that that the sets of rationalizable and confirmable outcomes are not independent since their

Subjective equilibrium imposes confirmability constraints on player beliefs but departs substantially from its alternatives in that players are not required to possess objective knowledge about the structure of the game. Note that Nash equilibrium, its extensions to Bayesian and correlated equilibria as well as to each of the alternative learning equilibria mentioned above, all assume:

1. Complete models: players know all structural details, including such elements as opponent identities, their payoffs, the informations structure, feasible actions, etc.

2. Correct common priors: players assign correct common priors to all unknown parameters.

3. Closed models: players assume that their opponents' understanding of the game is identical to their own.

Kalai and Lehrer argue that these assumptions are unrealistic and, therefore, of suspect predictive power. Thus, they are omitted in the subjective equilibrium analysis. Instead, each player is assumed to begin the game with her own subjective view of the personal consequences associated with her available strategies. More specifically, a player in the subjective equilibrium framework is endowed with her own "subjective environment response function," a function mapping from personal histories and actions to expected outcomes. In equilibrium, players maximize given their subjective environment response functions and observe outcomes with which they are consistent.

intersection always contains the set of Nash equilibria.

Since subjective equilibrium imposes objective knowledge neither of the game's structure nor of player rationality, it is possible to construct extreme cases in which virtually any outcome is supported as an equilibrium given the appropriate choice of player response functions. Thus, it is tempting to summarily dismiss this as a framework devoid of predictive content. To do so, however, would be to overlook a very useful analytical tool. Kalai and Lehrer provide a very general framework, a shell if you will, within which alternative sets of behavioral assumptions can be introduced and implications studied. One would expect such an ability to be valuable to those researching such issues as sophisticated learning (e.g., Fudenberg and Levine, 1998), bounded rationality (e.g., Rubinstein, 1998), and alternative utility formulations (e.g., Gul and Pesendorfer, 1998).

2. The model

A *causal-form* game is a 4-tuple $\Gamma = (N, A, \rightarrow, \pi)$ where $N \equiv \{1, \ldots, n\}$ is a finite set of players, $A \equiv \{A_i\}_N$ is the set of feasible actions for each player, $\rightarrow$ is the game's *causal structure* (described below), and $\pi : \Omega \rightarrow \mathbb{R}^n$ is a map from outcomes to payoffs. In order to simplify notational bookkeeping, assume that each active player moves exactly once so that $1, \ldots, m$ with $m \leq n$ indexes active players and $m + 1, \ldots, n$ indexes moves by nature. Let $\sigma(\cdot)$ denote the σ-algebra generated by a set.[3] Define $\mathcal{A}_i \equiv \sigma(A_i)$. The global set of outcomes is $\Omega \equiv \mathsf{X}_{i \in N} A_i$ with typical element ω. The product σ-algebra $\mathcal{F} \equiv \sigma(\mathsf{X}_{i \in N} \mathcal{A}_i)$ creates the measurable space $(\Omega, \mathcal{F})$. For each $i \in N$, let $a_i : \Omega \rightarrow A_i$ be the projection of ω onto A_i; note that an outcome

[3]For example, the Borel σ-algebra when the set is a topological space.

$\omega = (a_1(\omega), \ldots, a_n(\omega))$. Let $I_X : \Omega \to \{0,1\}$ be the indicator function that takes the value 1 if $\omega \in X \in \mathcal{F}$ and 0 otherwise.

If $(\Omega, \mathcal{F}, m)$ is a probability space, then $V \equiv \cup_N a_i \cup \pi$ is a set of random variables. The causal structure is assumed to be a primitive of the game relating these variables such that "$\to$" is a binary relation defined by $C \subset V \times V$ with the interpretation: if $v_i \to v_j$, then "v_i is a direct cause of v_j," and if $v_i \nrightarrow v_j$, then "v_i is not a direct cause of v_j." Dropping the player subscripts, an ordered subset $\{v_1, \ldots, v_k\} \subseteq V$ is a *causal chain*, if $v_j \to v_{j+1}$ for $j = 1, \ldots, k-1$. The chain is a *cycle* if $v_1 = v_k$. The relation $\to$ is *connected* if $\forall\ v, v' \in V$, there exists a chain containing v and v'. Assume that $\to$ is asymmetric, connected and acyclic.[4]

The *objective causal map of* Γ is defined as $(V, \to)$ which, from the preceding assumptions, is a directed, acyclic graph (with vertices V). Thus, it is possible to define the usual graph-theoretic objects. If $v, v' \in V$, then v' is a parent of v if $v' \to v$. Let P_v be the set of all v's parents. If $v, v' \in V$, then v' is a descendant of v if there exists a causal chain of the form $\{v, \ldots, v'\}$. Let D_v be the set of all descendants of v. An ordered subset $\{v_1, \ldots, v_k\} \subseteq V$ is a path if, for all $v_i, v_{i+1} \in \{v_1, \ldots, v_k\}$, either $v_i \to v_{i+1}$ or $v_{i+1} \to v_i$.

The causal structure of the game determines the information upon which players condition their behavior. Define $N_{\to i} \equiv \{j : a_j \in P_{a_i}\}$ as the set of players whose moves are direct causes of player i's action. Also, let $\sigma(v)$ for $v \in V$ denote the smallest σ-algebra with respect to which v is measurable.[5] If $Z \subseteq V$, then $\mathcal{H}_Z \equiv \sigma(Z)$ is said to be the *information algebra generated*

[4] Since $\to$ is acyclic, it is also irreflexive.

[5] So, $\sigma(a, a')$, where $a, a' \in A$, is shorthand for $\sigma(\sigma(a) \cup \sigma(a'))$, and so forth. Similarly, if $\mathcal{H}$ and $\mathcal{G}$ are sub-σ-algebras of $\mathcal{F}$, I write $(\mathcal{H}, \mathcal{G})$ rather

by Z. Player i's *information algebra* is $\mathcal{H}_i \equiv \sigma\left(a_j\right)_{j \in N_{\to i}}$, a sub-$\sigma$-algebra of $\mathcal{F}$. A *behavior strategy for player* i is an $\mathcal{H}_i$-measurable function $\sigma_i : \Omega \to \Sigma_i$ where Σ_i is the space of probability measures on $(A_i, \mathcal{F}_i)$. A *behavior strategy profile* is a $\sigma \in \Sigma \equiv \mathsf{X}_{i \in N} \Sigma_i$.[6] Denote the conditional probability given by $\sigma_i(\omega)$ to $X_i \in \mathcal{F}_i$ by $\sigma_i(X_i \,|\, \omega)$.

Lemma 1 *Let* $((\Psi_n, \mathcal{G}_n), n < \infty)$ *be a finite sequence of measurable spaces. Let* R_0 *be a probability measure on* $(\Psi_0, \mathcal{G}_0)$, *and for each* $n \geq 0$, *let* R_{n+1} *be a measurable function from* $(\Psi_0, \mathcal{G}_0) \times \ldots \times (\Psi_n, \mathcal{G}_n)$ *to the measurable space of probability measures on* $(\Psi_{n+1}, \mathcal{G}_{n+1})$. *Then there exists a probability space* $(\Omega, \mathcal{F}, m)$ *and random sequence* $v = (v_0, v_1, \ldots)$ *defined on that space such that the distribution of* v_0 *is* R_0 *and, for* $n \geq 0$, *a conditional distribution of* v_{n+1} *given* $\sigma(v_0, \ldots, v_n)$ *is given by* $\omega \mapsto R_{n+1}((v_0(\omega), \ldots, v_n(\omega)), \cdot)$. *For all* $X_n \in \mathcal{G}_0 \times \ldots \times \mathcal{G}_n$, *the distribution of* v *is uniquely determined by*

$$\rho[v \in X_n] = \int_{\Psi_0} \ldots \int_{\Psi_n} I_{X_n}(x_0, \ldots, x_n) R_n((x_0, \ldots, x_{n-1}), dx_n) \ldots R_0(dx_0). \tag{1}$$

Proof. This is a standard result, see Fristedt and Gray, 1997, p. 432. ■

Proposition 2 *Given a behavior strategy profile* $\sigma \in \Sigma$, *there exists a unique probability measure* m_σ *on* $(\Omega, \mathcal{F})$ *such that for all* $X \in \mathcal{F}$

$$m_\sigma(X) = \int_{V_1} \ldots \int_{V_n} I_X(\omega) \prod_{i \in N} \sigma_i\left(d\omega_i \,|\, (\omega_1, \ldots, \omega_n)\right). \tag{2}$$

---

than $\sigma(\mathcal{H} \cup \mathcal{G})$.

[6] The σ symbol serves double-duty, indicating both behavior strategies and the σ-algebras generated by sets. The intended meaning of σ will be clear from the context of its useage.

Proof. For all $a, a' \in V$, define "$\preceq$" such that $a \preceq a'$ if $a = a'$ or $a \neq a'$ and there exists a causal chain $\{a, \ldots, a'\}$. Let $(X)_{\preceq}$ denote a list of elements of $X \subset V$ ordered by $\preceq$. Define $\iota(k) \equiv i$ such that A_i is the k^{th} component of $(A_j : j \in N)_{\preceq}$. Let $p_1 \equiv \sigma_{\iota(1)}$. For each $r \geq 1$, let $p_{r+1} : \Omega \to \Sigma_{\iota(r+1)}$ be $\mathcal{H}_{\iota(r+1)}$-measurable where, for all $X \in \mathcal{H}_{\iota(r+1)},$ $p_{r+1}((\omega), X) \equiv \sigma_{\iota(r+1)}(X \mid \omega)$. Note that, by the construction of $\preceq$, $P_{v_{r+1}} \subseteq \left\{a_{\iota(1)}, \ldots, a_{\iota(r)}\right\}$. Therefore p_{r+1}, being $\mathcal{H}_{\iota(r+1)}$-measurable is also $\left(A_{\iota(1)}, \mathcal{F}_{\iota(1)}\right) \times \ldots \times \left(A_{\iota(r)}, \mathcal{F}_{\iota(r)}\right)$-measurable. By Lemma 1, for all $X \in \sigma\left(\mathcal{A}_{\iota(1)} \times \ldots \times \mathcal{A}_{\iota(n)}\right)$, the distribution of $a \equiv \left(a_{\iota(1)}, \ldots, a_{\iota(n)}\right)$ is uniquely determined by the probability measure p where, for all $X \in \mathcal{F}$,

$$p(X) = \underbrace{\int \ldots \int}_{(A_j : j \in N)_{\preceq}} I_X(x)\, p_{\iota(n)}\left(\left(x_{\iota(1)}, \ldots, x_{\iota(n-1)}\right), dx_{\iota(n)}\right) \times \ldots \times p_{\iota(1)}\left(dx_{\iota(1)}\right). \tag{3}$$

Noting that $\mathcal{F} = \sigma\left(\mathcal{A}_{\iota(1)} \times \ldots \times \mathcal{A}_{\iota(n)}\right)$, $m_\sigma(X) = p(X)$, and substituting back the σ_i's gives the desired result. ■

3. Preliminary Results

Notice that the set of functions V are vertices in the graph $(V, \to)$ and are random variables on the probability space $(\Omega, \mathcal{F}, m_\sigma)$. The following result describes how these objects are related. First, v_i is said to *activate* the path $\{v_1, \ldots, v_k\}$ if it is interior $(v_i \in \{v_1, \ldots, v_k\}, v_1 \neq v_i \neq v_k)$ and $v_{i-1} \to v_i \leftarrow v_{i+1}$.

Definition 3 Let $X, Y, Z \subset V$ be disjoint. Then we say the causal effect of X on Y is *blocked* by Z, denoted $(X \nrightarrow Y \mid Z)$, if for all $(x, y) \in X \times Y$ and all paths $\{x, \ldots, y\}$, there exists a v such that:

1. v activates $\{x, \ldots, y\}$ and $(v \cup D_v) \cap Z = \emptyset$, or
2. v does not activate $\{x, \ldots, y\}$ and $v \in Z$.

Although its graph-theoretic orientation may be somewhat unfamiliar, the intuition behind this definition is fairly straightforward. If v_i and v_j are causally related, either as elements in a causal chain or through a set of shared causes, then knowing the value of one provides relevant information with respect to the value of the other. However, this is no longer true once the value of intermediate variables or of common causes is revealed.

Assume that players begin the game with knowledge of player identities and their action sets. That is, players know V_i for all $i \in N$. Players also know whose moves they will, themselves, observe at the time of their move (so, player i knows$N_{\rightarrow i}$ and, by implication, $\mathcal{H}_i$). Players also know their own component of the payoff function. This leaves uncertainty regarding $\rightarrow$, the game's causal structure, and σ, the behavior of opponents.

Let $m_\sigma (F \mid \mathcal{G})$ denote the conditional probability of $F \in \mathcal{F}$ given a sub-σ-algebra $\mathcal{G}$ of $\mathcal{F}$. Two sub-σ-algebras $\mathcal{G}_1, \mathcal{G}_2$ are stochastically independent given a third $\mathcal{G}_3$ if, for all $G_i \in \mathcal{G}_i$, $m_\sigma (G_1 \cap G_2 \mid \mathcal{G}_3) = m_\sigma (G_1 \mid \mathcal{G}_3)\, m_\sigma (G_2 \mid \mathcal{G}_3)$; denote this relationship $(\mathcal{G}_1 \perp \mathcal{G}_2 \mid \mathcal{G}_3)_\sigma$.

Proposition 4 *For all* $X, Y, Z \subset V$ *and all* $\sigma \in \Sigma$, $(X \nrightarrow Y \mid Z) \Rightarrow (\mathcal{G}_X \perp \mathcal{G}_Y \mid \mathcal{G}_Z)_\sigma$.

Proof. See Pearl (1988). ■

In order to separate the problem of causal uncertainty from that of behavioral uncertainty, assume for the moment that players know m_σ, the actual distribution over outcomes, even though they do not know σ itself.[7] Let $\mathbf{C}$ be the (finite) set of all possible causal relations on $V \times V$ with typical element C_r defining the causal map $(V, \rightarrow_r)$.

For each $C_r \in \mathbf{C}$, let $N_{i|r} \equiv \{j : (v_j, v_i) \in C_r\}$ be the set of players whose moves are observed by player i in $(V, \rightarrow_r)$. Then, player i's *information algebra under* C_r is $\mathcal{H}_{i|r} \equiv \sigma\left(v_j\right)_{j \in N_{i|r}}$. Player i's set of C_r-consistent behavior strategies is the space of $\mathcal{H}_{i|r}$-measurable functions from Ω to Σ_i. Denote this space $\Sigma_{i|r}$ with typical element $\sigma_{i|r}$. Then, the space of C_r-consistent behavior strategy profiles is $\Sigma_r \equiv \mathsf{X}_{i \in N} \Sigma_{i|r}$ with typical element σ_r. If players know the frequency over outcomes m_σ, then the result of Proposition 2 provides an obvious definition for causal consistency.

Definition 5 For all $(C_q, C_r) \in C \times C$, C_q is said to be a *minimal set of* C_r*-consistent causal relations*, denoted $C_q \sim C_r$, if for all $\sigma_r \in \Sigma_r, X \in F$ there exists a $\sigma_q \in \Sigma_q$ satisfying $m_{\sigma_q}(X) = m_{\sigma_q}(X)$, where

$$m_{\sigma_q}(X) \equiv \int_{V_1} \cdots \int_{V_n} I_X(\omega) \prod_{i \in N} \sigma_{i|q}\left(d\omega_i \mid (\omega_1, \ldots, \omega_n)\right), \tag{4}$$

and removal of any $(v, v') \in C_r$ causes this condition to fail. Let $\bar{C}$ be the set of all C_q such that $C_q \sim C$.

In other words, if $C_r \in \bar{\mathbf{C}}$, then $(V, \rightarrow_r)$ is a causal map for which there exists a $\sigma_r \in \Sigma_r$ that induces a probability measure m_{σ_r} on outcomes such

[7] Such knowledge (i.e., of outcome frequency versus actual strategies) might well arise following a learning process in which players observe play for an extended period.

that $m_{\sigma_r} = m_\sigma$. Note that, when $\rightarrow_r$ is consistent with $\rightarrow$, all of the conditional independencies implied by $\rightarrow_r$ are indeed present in m_σ. The idea is that a player with causal beliefs $C_r \in \bar{\mathbf{C}}$ can always construct a σ_r consistent with m_σ, the actual distribution over outcomes. Thus, for all $C_r \in \bar{\mathbf{C}}$, the hypothesis $(C = C_r)$ can never be disconfirmed solely by knowledge about the actual distribution over outcomes. The last part of the definition eliminates maps with superfluous assertions of dependency. One obvious question is just how large $\bar{\mathbf{C}}$ might be.

Definition 6 The ordered pair $(v_i, v_j) \in V \times V$ is said to be an *undirected relationship* under C_r if and only if $(v_i, v_j) \in C_r$ or $(v_j, v_i) \in C_r$.

Let $R_r \subseteq V \times V$ denote the set of all undirected relationships under C_r.

Definition 7 Two undirected relationships $(a_1, a_2), (b_1, b_2) \in R_r$ are said to contain *common causes* if and only if they each contain a common consequence; that is

1. $\exists\ i, j \in \{1, 2\}$ such that $a_i = b_j$ and
2. (a_{-i}, a_i) , $(b_{-i}, b_i)\, C_r$

The relationships are said to exhibit *causal independence* if and only if the above conditions are met and the common causes are directly and indirectly unrelated; that is $(v_{-i}, b_{-i}) \notin R_r$ and $\nexists\ z \in V$ such that the sequences $z, \ldots, v_{-i}$ and $z, \ldots, b_{-i}$ are causal chains.

Let $IC_r \subseteq C_r$ denote the set of all relationships exhibiting causal independence.

Proposition 8 *Suppose* $(C_q, C_r) \in \mathbf{C} \times \mathbf{C}$, *then*

$$C_q \sim C_r \Leftrightarrow (R_g = R_r \text{ and } IC_g = IC_r)\,. \tag{5}$$

Proof. See Verma and Pearl, 1990. ■

First, it is readily apparent from this proposition that $\bar{\mathbf{C}}$ is not in general a singleton set. Second, the proposition allows us to specify – by simple inspection of the graph $(V, \rightarrow)$ – the set of all consistent causal maps, where consistency is defined as any map implying the same stochastic independencies. The proposition says that causal maps are consistent if they have the same edges, regardless of direction, and the same direction on edges that emanate from independent common causes. Thus, the only time causal directionality is pinned down is when two mutually independent elements of V are direct causes of a third.

Proposition 9 *The relation "$\sim$" is an equivalence relation.*

Proof.

1. *Reflexivity* Clearly, for all $C_r \in \mathbf{C}$, $C_r \sim C_r$,
2. *Transitivity* Suppose $C_q \sim C_r$ and $C_r \sim C_s$. From Proposition 8, C_q, C_r, and C_s all have equivalent sets of undirected and causally independent relationships. Thus, again invoking Proposition 8, $C_q \sim C_s$.
3. *Symmetry* $C_q \sim C_r \Rightarrow C_r \sim C_q$ also follows directly from Proposition 8.

■

Up to this point, we have considered a fairly strong form of causal consistency, namely the one that holds under all possible behaviors (i.e., for all of Σ). Thus, the makeup of $\bar{\mathbf{C}}$ depends only upon Γ, not player behavior. However, there may be instances when we are interested in consistency for a particular σ. For example, in equilibrium the set of casual maps in the support of player beliefs should be those consistent with a particular m_σ.

Recall that $\mathcal{H}_v$ is the algebra generated by the random variable v. A player's action variable a_j is said be *σ-relevant to i*, denoted $a_j \rightarrow_\sigma a_i$, if $j \in N_{\rightarrow i}$ and there exist $H, H' \in \mathcal{H}_{a_j}$, such that $m_\sigma(H), m_\sigma(H') > 0$ and for all $\omega \in H$, $\omega' \in H'$ $\sigma_i(\omega) \neq \sigma_i(\omega')$. This condition is met, for example, when at least two actions that player j chooses with positive probability under σ_j cause player i to respond differently under σ_i. The *σ-identified causal structure* is $C_\sigma \equiv \{(v_j, v_i) \in C \mid v_j \rightarrow_\sigma v_i\}$ and forms a subgraph of $(V, \rightarrow)$, denoted $(V, \rightarrow_\sigma)$.

Differences between C and C_σ arise when $j \in N_{\rightarrow i}$, but a_j is not *σ-relevant to i*. This situation can occur either when player j implements a pure strategy or when j mixes on actions that player i chooses to ignore. In either case, (a_j, a_i) is removed from C. Thus, $C = \cup_\Sigma C_\sigma$. Define the set of *σ-consistent causal relations*, denoted $\bar{\mathbf{C}}_\sigma$, as those $C_r \in \mathbf{C}$ such that $C_r \sim C_\sigma$. When C_σ is known, $\bar{\mathbf{C}}_\sigma$ can be constructed directly using Proposition 8. Suppose, however, that players only know the distribution m_σ.

Proposition 10 *Given a game Γ and a probability distribution over outcomes m_σ, $C_r \sim C_\sigma$ if and only if for all $v \in V$, $v' \notin D_v$, $(\sigma(v) \perp \sigma(v') \mid \sigma(P_v))_\sigma$ and this condition fails for all proper subsets of P_v.*

Proof. See Pearl, 1988. ■

The proposition says that, given some σ, any $(V, \rightarrow_r)$ constructed such that each action variable is stochastically independent of all of its non-descendants given its parent set contains a minimal set of consistent causal relations. Note the empirical implication of this proposition: if one can estimate m_σ, one can then construct an estimate of $\bar{\mathbf{C}}_\sigma$.

4. Causal Ambiguity

An interesting implication of the preceding results is that they allow us to specify the degree of possible confusion over game structure either as a direct function of the game itself or indirectly from a specific description of player behavior. Define a game's *intrinsic degree of causal ambiguity* as $|\bar{\mathbf{C}}|$, the set cardinality of $\bar{\mathbf{C}}$ (i.e., the number of relations consistent with the true causal structure of the game). From Proposition 8, we know that the structure of some games is inherently transparent, while that of others is fundamentally opaque. Similarly, define $\left|\bar{\mathbf{C}}_\sigma\right|$ to be the *degree of causal ambiguity under* σ.

Proposition 11 *If $\sigma \in \Sigma$ is a strategy profile such that for all $i \in N, j \in N_{\rightarrow i}$, a_j is σ-relevant to i, then $\left|\bar{\mathbf{C}}\right| = \left|\bar{\mathbf{C}}_\sigma\right|$.*

Proof. Clearly, $C_\sigma \subseteq C$. Suppose $(v_j, v_i) \in C$ but $(v_j, v_i) \notin C_\sigma$. Then, there exists a $j \in N_{\rightarrow i}$ such that a_j is not *σ-relevant to i*, immediately violating the premise of the proposition. ■

Thus, an important case in which the set of C-consistent maps is fully identified is when σ is an equilibrium in which: i) all players mix (as in a game with trembles); and, ii) a player is never asked to observe payoff-irrelevant information (as in any parsimoniously constructed model).

It may be tempting to conclude that $\left|\bar{\mathbf{C}}\right| \leq \left|\bar{\mathbf{C}}_\sigma\right|$; that is, causal ambiguity is greater when a particular behavior profile hides a certain number of causal connections. Unfortunately, such intuition is incorrect – causal ambiguity is not monotonically related to the number of edges revealed in $(V, \rightarrow)$.

Proposition 12 *Assume $C_q, C_r \in \mathbf{C}$. Then*

$$C_q \subset C_r \nRightarrow \left|\bar{\mathbf{C}}_q\right| > \left|\bar{\mathbf{C}}_q\right| \vee \left|\bar{\mathbf{C}}_q\right| < \left|\bar{\mathbf{C}}_q\right|. \tag{6}$$

Proof. Suppose $(v_1, v_2) \in C_r \backslash C_q$. Consider $C_{r1} \equiv C_r \backslash (v_1, v_2)$. If $(v_1, v_2) \in IC_r$, then from Proposition 8, $\left|\bar{\mathbf{C}}_{r1}\right| \geq \left|\bar{\mathbf{C}}_r\right|$: either i) all of the relationships that exhibit causal independence with (v_1, v_2) also exhibit causal independence with some other relation in IC_r, implying $\left|\bar{\mathbf{C}}_{r1}\right| = \left|\bar{\mathbf{C}}_r\right|$; or, ii) at least one relationship must be removed from IC_r when (v_1, v_2) is removed, implying $\left|\bar{\mathbf{C}}_{r1}\right| > \left|\bar{\mathbf{C}}_r\right|$. On the other hand, if $(v_1, v_2) \notin IC_r$, then the direction of (v_1, v_2) is ambiguous, so $\left|\bar{\mathbf{C}}_{r1}\right| < \left|\bar{\mathbf{C}}_r\right|$. For the desired result, simply continue the removal of one edge at a time until C_r is reduced to C_q. ■

Proposition 13 *Given $\Gamma = (N, A, \rightarrow, \pi)$ is a causal-form game,*

1. *if $\rightarrow$ describes a simultaneous move game, then $\left|\bar{\mathbf{C}}\right| = 1$,*
2. *if $\rightarrow$ describes a game of perfect information,*
   then $\left|\bar{\mathbf{C}}\right| = \max_{C_r \in \mathbf{C}} \left|\bar{\mathbf{C}}_r\right| = 2^n$.

Proof. Part 1 is true because, in a simultaneous move game, there are no causal dependencies that create ambiguity; all edges are of the form (a_i, π) and, as such, belong to IC. Part 2 is true by just the opposite reasoning. In a game of perfect information, the graph $(V, \rightarrow)$ is fully connected. As a result, the direction of every edge is ambiguous, resulting in the maximum number of consistent graphs (2^n). ■

Proposition 14 *Let $C_q, C_r \in \mathbf{C}$ where C_r is the game formed by adding γ new observation relationships to C_q (so, $|C_r \backslash C_q| = \gamma > 0$). Then, there exists some threshold number of observations k such that for all $\gamma \geq k$, $|\bar{\mathbf{C}}_r| > |\bar{\mathbf{C}}_q|$.*

Proof. Let $C^* \in \mathbf{C}$ be any fully connected graph. Then, the result is guaranteed for $\overline{k} \equiv |\bar{\mathbf{C}}_r| - |\bar{\mathbf{C}}_q|$ such that $C_r = C^*$. In general, first, augment C_q by adding the γ relevant observation relations required to generate a new game C_q' that contains the maximum number of causally independent relations given the original causal structure. This results in the minimum level of ambiguity that can be obtained by adding relations. Next, add the minimum number, γ', of new relations to create C_q'' such that the maximum number of causally independent relations is eliminated from C_q' subject to $|\bar{\mathbf{C}}_q''| \leq |\bar{\mathbf{C}}_q'|$. Now, adding any single relation to C_q'' creates a game C_r for which $|\bar{\mathbf{C}}_r| > |\bar{\mathbf{C}}_q|$. Therefore, $k = \gamma + \gamma'$ provides the result under this algorithm. ■

5. Causal Equilibrium

It is now possible to consider causal beliefs in equilibrium. Let $\Delta(\cdot)$ denote the space of probability measures on a set. Then $\mu_i \in \Delta(\mathbf{C})$ represents player i's beliefs regarding the true causal nature of the game; that is, $\mu_i(C_r)$ is i's subjective assessment of the proposition $(C_r = C)$. Let $\rho_{i|r} \in \Delta\left(\Sigma_{-i|r}\right)$

be the function identifying player i's beliefs about opponent strategies under each C_r. Player i's expected payoff given his beliefs and strategy is

$$u_i\left(\sigma_i, \mu_i, \rho_i\right) \equiv \sum_{\mathbf{C}} \mu_i\left(C_r\right) \int_{\Sigma_{-i|r}} \int_{\Omega} \pi_i m_{\left(\sigma_i, \sigma_{-i|r}\right)}(d\omega)\, \rho_{i|r}\left(d\sigma_{-i|r}\right) \qquad (7)$$

Definition 15 A *causal equilibrium* is a σ and beliefs $\left(\mu_i, \rho_i\right)_N$ such that

1. $\forall\ \sigma_i' \in \Sigma_i,\ u_i\left(\sigma_i, \mu_i, \rho_i\right) \geq u_i\left(\sigma_i', \mu_i, \rho_i\right),$

2. $\forall\ C_r \in \text{support}\left(\mu_i\right),\ \sigma_{-i|r} \in \text{support}\left(\rho_{i|r}\right) \Rightarrow m_{\left(\sigma_i, \sigma_{-i|r}\right)} = m_\sigma.$

The first condition requires players to play best responses to their beliefs. The second item requires the players' causal beliefs to be consistent with the actual distribution over outcomes as well as with knowledge of his own information algebra. The last says that beliefs must be consistent with the actual distribution over outcomes. Note that, for all $\sigma_r, \sigma_r' \in \Sigma_r$ satisfying definition 5, the expected value of π_i is the same (since, by definition, they induce probability measures over outcomes equivalent to m_σ). The margin for error in a causal equilibrium is a misunderstanding of who sees what before moving as well as incorrect assessments of player behavior off the equilibrium path.

Definition 16 A *Nash equilibrium* is a σ and $\left(\mu_i, \rho_i\right)_N$ such that

1. $\forall\ \sigma_i' \in \Sigma_i,\ u_i\left(\sigma_i, \mu_i, \rho_i\right) \geq u_i\left(\sigma_i', \mu_i, \rho_i\right),$

2. $\forall\ C_r \in \text{support}\left(\mu_i\right),\ \sigma_i' \in \Sigma_i,$
   $\sigma_{-i|r} \in \text{support}\left(\rho_{i|r}\right) \Rightarrow m_{\left(\sigma_i', \sigma_{-i|r}\right)} = m_{\left(\sigma_i', \sigma_{-i}\right)}.$

The requirements are that: i) players play optimal strategies given their beliefs; and, ii) their assessments regarding the consequences of their feasible strategy choices are correct (the causal relations in the support of μ_i must hold for all σ_i').

Proposition 17 *If* $(\sigma, (\mu_i, \rho_i)_N)$ *is a Nash equilibrium, then* $\forall i \in N$, $\mu_i \left(\left\{ C_r \in \bar{\mathbf{C}}_\sigma \mid P_{v_i|r} = P_{v_i} \right\} \right) = 1$.

Proof. This is just definition chasing. $P_{v_i|r} = P_{v_i}$ by the assumption that each player knows their own information structure. Suppose $C_r \notin \bar{\mathbf{C}}_\sigma$. Then, by Definition 5, some $\left(\sigma_i, \sigma_{-i|r}\right)$ cannot be constructed such that $m_{\left(\sigma_i, \sigma_{-i|r}\right)} = m_\sigma$. ■

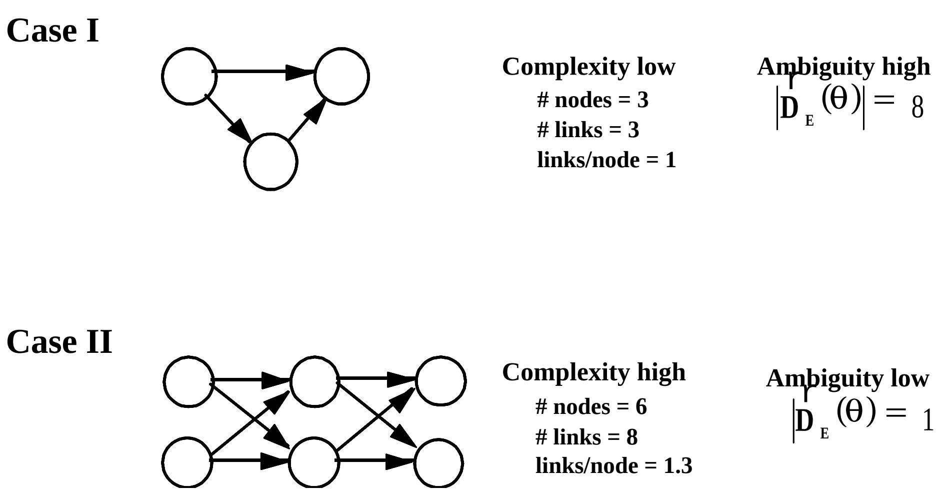
Let $CE, NE \subseteq \Sigma$ denote the set of profiles capable of supporting a causal or Nash equilibrium, respectively. Suppose μ_i^{CE} is part of a causal equilibrium for some player i and μ_i^{NE} is part of a Nash equilibrium for the same player in the same game. Since any causal relation receiving positive weight under μ_i^{NE} must have at least the same number of edges as any counterpart under μ_i^{CE}, it is tempting to conclude that $\left|\text{support}\left(\mu_i^{CE}\right)\right| \geq \left|\text{support}\left(\mu_i^{NE}\right)\right|$; that is, causal ambiguity in a causal equilibrium (clearly a "looser" concept than Nash) is greater than that in a Nash equilibrium. Unfortunately, as we know from Proposition 12, such reasoning does not generally hold.

6. Complexity Versus Ambiguity

Reed and DeFillippi (1990) forward the "common sense" hypothesis that causal ambiguity is exponentially increasing in the level of environmental complexity. In this section, I present a specific analytical measure of causal ambiguity, $|\bar{\mathbf{C}}|$, and demonstrate that this common sense hypothesis fails.

First, consider several potential candidates as measures of environmental complexity. Some that come immediately to mind are the number of players in the game (nodes), the number of player interactions (links between variables), and the number observations per player (links per variable). The intuition is that as the system becomes more complex along these dimensions, the difficulty of causal assessment increases.

As demonstrated in the following figure, the relationship between environmental complexity and causal ambiguity is not exponential, monotonic, or even non-decreasing. The relationship is subtle and depends not only upon the complexity of the system, but also upon the specific nature of its causal connections. In a sense, the Reed and DeFillippi hypothesis is correct: as we know from Proposition 14, beyond a certain level of complexity, causal ambiguity will increase (with the unique maximum level attained in the fully connected system).



complexity versus causal ambiguity

Carol Simon Hall, University of Rochester, Rochester, NY 14627;
ryall@ssb.rochester.edu.

References

[1] Abreu, D., D. Pearce, and E. Stacchetti (1990): "Toward a Theory of Discounted Repeated Games with Imperfect Monitoring." *Econometrica,* 58, 1041-1063.

[2] Battigalli, P., and D. Guatoli (1988): "Conjectural Equilibria and Rationalizability in a Macroeconomic Game with Incomplete Information," mimeo. Instituto Economia Politica, Milan.

[3] Blume, L. E., and D. Easley (1992): "Rational Expectations and Rational Learning," working paper. Cornell University, Ithica .

[4] Brandenberger, A. (1993): "Strategic and Structural Uncertainty in Games,"working paper. Harvard University, Boston .

[5] Chickering, D. M. (1996): "Learning Bayesian Networks from Data." Ph. D. Dissertation, University of California, Los Angeles.

[6] Einhorn, H. J., and R. M. Hogarth (1986): "Decision Making Under Ambiguity," *Journal of Business,* 59, S225-S250.

[7] Fristedt, B., and L. Gray (1997): *A Modern Approach to Probability Theory.* Boston: Birkhauser.

[8] Fudenberg, D., and D. K. Levine (1993): "Self-confirming Equilibrium," *Econometrica,* 61, 523-45.

[9] —— (1998): *Theory of Learning in Games.* Cambridge: The MIT Press.

[10] Geanakoplos, J. (1994): "Common Knowledge," in *Handbook of Game Theory*, edited by R. J. Auman and S. Hart, 1437-1497: Elsevier Science B. V.

[11] Granger, C. (1969): "Investigating Causal Relations by Econometric Models and Cross-Spectral Methods," in *Econometrica*, 37, 424-38.

[12] Gul, F., and W. Pesendorfer (1998): "Dynamic Inconsistency and Self-Control," mimeo: Princeton University.

[13] Harsanyi, J. (1967-68): "Games with Incomplete Information Played by 'Bayesian' Players," *Management Science* 8, Parts I-III 159-82,320-334,486-502.

[14] Howitt, P., and R. P. McAfee (1992): "Animal Spirits," *The American Economic Review,* 82, 493-507.

[15] Kalai, E., and E. Lehrer (1993): "Subjective Equilibrium in Repeated Games," *Econometrica,* 61, 1231-1240.

[16] —— (1995): "Subjective Games and Equilibria," *Games and Economic Behavior,* 8, 123-63.

[17] Lippman, S. A., and R. P. Rumelt (1988): "Uncertain Imitability: An Analysis of Interfirm Differences in Efficiency Under Competition," *The Bell Journal of Economics,* 13, 418-38.

[18] Milgrom, P., and J. Roberts (1990): "Rationalizability. Learning, and Equilibrium in Games with Strategic Complementarities," *Econometrica,* 58, 1255-77.

[19] Nachbar, J. H. (1996): "Prediction, Optimization and Rational Learning in Games,"Working paper. Washington University, St. Louis.

[20] Pearce, D. G. (1984): "Rationalizable Strategic Behavior and the Problem of Perfection," *Econometrica,* 52, 1029-1050.

[21] Pearl, J. (1988): *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference.* 1 ed. San Francisco: Morgan Kaufmann.

[22] Reed, R., and R. DeFillippi (1990): "Causal Ambiguity, Barriers to Imitation, and Sustainable Competitive Advantage," *Academy of Management Review,* 15, 88-102.

[23] Rubinstein, A., and A. Wolinsky (1990): "Rationalizable Conjectural Equilibrium: Between Nash and Rationalizability," mimeo: 19-90. Tel-Aviv University, Ramat Aviv.

[24] Sims, C. (1972): "Money, Income, and causality," *American Economic Review,* 62, 540-52.

[25] Tversky, A., and D. Kahneman (1977): "Causal Schemata in Judgements Under Uncertainty." In *Progress in Social Psychology*, edited by M. Fishbein. Hillsdale, NJ: Lawrence Erlbaum.

[26] Verma, T. S., and J. Pearl (1991): "Equivalence and Synthesis of Causal Models," *Uncertainty in Artificial Intelligence,* 6, 255-68.