Common Knowledge in Simultaneous-Move Games

Paul Weirich

Philosophy Department

University of Missouri

Columbia, MO 65211

phone: 573-882-6760

e-mail: weirichp@missouri.edu

draft 2-09

1. Epistemic Logic

Decision theory and game theory evaluate reasoning concerning action. Such reasoning may involve reasoning about knowledge. For example, a player in a game may reason about the knowledge of other players to infer their acts and adopt a best response.

Epistemic logic typically advances principles of reasoning for ideal agents. The ideal agents have no cognitive limits. A theory of reasoning uses idealizations to control for explanatory factors. The theory may later dispense with the idealizations to obtain a more realistic theory of reasoning by humans. Some computer scientists construct artificial reasoners as similar as possible to ideal reasoners. A theory of reasoning for ideal agents assists the design of artificial reasoners.

Studying strategic reasoning clarifies assumptions and solidifies conclusions about behavior in games. This paper uses familiar formal systems such as first-order logic and set theory. It formulates premisses that support standard conclusions about behavior in games. It uses principles of reasoning that, for selected games of strategy, support a player's participation in a Nash equilibrium. The principles come from epistemic logic and concern common knowledge. The sections explain common knowledge, its origin, and its consequences for action. They describe games of strategy and their Nash equilibria. Finally, they show how players' reasoning leads them to participation in a Nash equilibrium. Principles of epistemic logic specify players' knowledge about other players' strategies and explain how that knowledge supports their participation in a Nash equilibrium of the game.

One method of analyzing reasoning examines the structure of reasons. Logic, for example, organizes the reasons for an argument's conclusion. The justification of an act may explain the beliefs and desires that furnish reasons for the act.[1]

A rational person may be unaware of or not appreciate reasons he possesses. Consequently, his reasons for a belief may not yield the belief. He may fail to believe complex mathematical truths. A valid argument with premisses that he believes may not produce a belief in the conclusion because he does not formulate the argument. The principle of deductive closure expresses an ideal and not a requirement of rationality. Similarly, a rational person may fail to comply with the KK principle. Knowing puts one in position to know that one knows, but may not yield that knowledge. If one is unaware of reasons that put one in position to know, knowledge may not arise.[2]

An ideal reasoner is aware of all reasons she has and responds to them. An account of reasoning for ideal agents may treat the structure of reasons. However, it may instead treat directly the reasoning of ideal agents without elaborating an account of the structure of their reasons. For simplicity, this paper forgoes an account of reasons. Although an account of the structure of reasons applies to reasons that both humans and ideal agents possess, it is simpler to treat only the reasoning of ideal agents.

---

[1] Some accounts of knowledge claim that standards of knowledge are context-sensitive. According to these accounts, whether a person knows a proposition depends on the agent's plan of action. A traveler may know that a plane about to depart from an airport gate is going to Chicago because the departure monitors say so. However, the traveler may not know the plane's destination if he is about to board the plane and remembers announcements about the possibility of gate changes. In that case, his knowing that the plane is going to Chicago may require confirmation from the ticket agent. The standards for knowledge may rise in response to its practical importance. If contextualism is right, then an account of knowledge requires attention to reasons supporting acts and not just reasons supporting beliefs.

[2] Bergmann (2006: Chap. 1) discusses knowledge and awareness of reasons.

Although strategic reasoning in games may involve belief rather than knowledge, for convenience, this paper treats ideal games in which players' beliefs are accurate and warranted so that they constitute knowledge.

2. Common Knowledge

Common knowledge exists in a population and in its technical sense is more than knowledge shared throughout the population. A population has common knowledge of a proposition if and only if in the population everyone knows that proposition, everyone knows that everyone knows that proposition, everyone knows that everyone knows that proposition, and so on ad infinitum.

To couch this definition in compact notation, one may introduce notation for mutual knowledge and then use it to define common knowledge. People in a population have mutual knowledge that p if and only if all know that p. Let $(EK)p$ stand for mutual knowledge that p. Let $(EK)^2p$ abbreviate $(EK)(EK)p$, let $(EK)^n p$ abbreviate a string of n (EK)'s followed by p, and let $(EK)^\infty p$ abbreviate for all n, $(EK)^n p$. In the population, there is common knowledge that p if and only if $(EK)^\infty p$.

This familiar account of common knowledge leaves open arrangement of quantifiers and so whether $(EK)^2p$ stands for $\forall x K_x \forall y K_y p$ or $\forall x \forall y K_x K_y p$. The usual disambiguation takes the second interpretation. Hence an ideal person's knowing that all know that p is equivalent to the person's knowing a conjunction in which each conjunct expresses a member of the population's knowledge that p. The person may know the conjunction without knowing that it covers every member of the population.

In a familiar story illustrating the relevance of common knowledge to action, two women, Alice and Betty, strangers to each other, are traveling in a compartment of a train. Each woman's face bears the grime of travel. She sees that the other woman's face is dirty but does not see that her own face is dirty. Their compartment tickets issue turns for the lavatory. As they know, Alice's turn is first and Betty's turn is second. Each knows that the other cleans her face if and only if she knows it is dirty. Neither woman knows her face is dirty, so neither woman cleans her face.

The train's conductor enters the compartment to collect tickets and, to be helpful, mentions that someone has a dirty face. Each woman already knew this, but now knows that each knows this, knows that each know this, and so on. The two women acquire common knowledge that at least one has a dirty face.[3]

After the conductor's announcement, Alice forgoes her turn to use the lavatory. She thinks that Betty's dirty face prompted the conductor's announcement. Betty observes that Alice does not clean her face. She infers that Alice does not clean her face because Alice sees that Betty's face is dirty. So Betty uses her turn to wash her face.

To add precision to this argument that Betty washes her face, I assume some principles concerning the structure of knowledge. The principles come from applications of modal logic to epistemic possibility. A modal account of knowledge defines knowledge in terms of epistemic possibility. A person knows a proposition p, that is, Kp, if and only if p is true in every world epistemically possible for the person. This account of knowledge relies on traditional

[3] This is a version of an example in Binmore and Brandenburger (1990).

epistemology for an explanation of epistemic possibility. Assuming that an S5 modal system governs epistemic possibility, these axioms about knowledge hold.[4]

(K0) If *p* is necessarily true, then *Kp*.

(K1) *K*(*p* & *q*) if and only if (*Kp* & *Kq*).

(K2) If *Kp*, then *p*.

(K3) If *Kp*, then *KKp*.

(K4) If ~*K*~*Kp*, then *Kp*.

Idealizations ground principles governing the structure of knowledge such as (K0)–(K4). These principles about knowledge assume that agents are ideal and have no cognitive limits. Moreover, ideal agents know the principles and use them to make inferences about other ideal agents' knowledge.

To analyze the argument of the example, I use a consequence of K1. According to K1, *K*(*p* & *q*) if and only if (*Kp* & *Kq*). If *p* entails *q*, then *p* & *q* is equivalent to *p*. So *Kp* is equivalent to *K*(*p* & *q*). If *K*(*p* & *q*), then *Kq*, by K1. Hence if *p* entails *q*, then *Kp* yields *Kq*. According to K1, an ideal agent knows any proposition entailed by any proposition he knows. Moreover, an ideal agent knows any proposition entailed by any conjunction of propositions he knows.[5]

I also use K2 and K1 to derive the subsidiary principle that if an agent knows that someone knows that *p*, then the agent knows that *p*, that is, $K_j K_i p \rightarrow K_j p$. For suppose that agent *j* knows

---

[4] See, for example, Binmore and Brandenburger (1990: 108). These principles of knowledge apply to common knowledge also. Although S5 is a system of propositional modal logic, the propositions involved may be generalizations. The system lacks only quantification into a modal formula.

[5] For simplicity, I assume that a proposition is a set of possible worlds. Equivalent propositions are the same set of possible worlds. Hence if *p* and *q* are equivalent, knowledge of *p* amounts to knowledge of *q*. This account of the objects of knowledge suits ideal agents.

that $K_i p$. By K2, the proposition he knows entails p. So by K1, $K_j p$. Granting contextualism about knowledge, this inference assumes as an idealization that agents are in contexts that yield the same standards of knowledge for each.

Next, I introduce some terminology to obtain compact expressions of propositions that appear in the example's argument. Let K_A express knowledge for Alice, and let K_B express knowledge for Betty. Let DA express that Alice's face is dirty, and let DB express that Betty's face is dirty. Before the conductor speaks, $K_A DB$ & $K_B DA$, and furthermore ~$K_A DA$ & ~$K_B DB$. Let CA express that Alice cleans her face, and let CB express that Betty cleans her face. Because the women know about their dispositions to wash, they know that $CA \leftrightarrow K_A DA$ and that $CB \leftrightarrow K_B DB$. From the women's knowledge and dispositions to wash, it follows that ~CA & ~CB.

Let S be the disjunction DA v DB. After the conductor's announcement, the two women have common knowledge that S. Using terminology introduced to express common knowledge, $(EK)^\infty S$. From this common knowledge, it follows that $(EK)^2 S$. Because A and B form the relevant population, $(EK)^2 S$ if and only if $K_A(K_A S$ & $K_B S)$ & $K_B(K_A S$ & $K_B S)$.

The following argument displays in detail the example's inferences about Betty's knowledge and acts.

1. $(EK)^\infty S$	A result of the conductor's announcement
2. $CA \leftrightarrow K_A DA$	Dispositions to wash
3. K_B~$K_A DA$	2, Betty's observation that Alice does not clean her face, K1
4. $(EK)^2 S$	1, the definition of common knowledge
5. $K_B K_A S$	4, the definition of mutual knowledge, K1
6. $K_B(K_A S$ & ~$K_A DA)$	3, 5, K1

7. $K_B[(K_AS \;\&\; {\sim}K_ADA) \rightarrow K_ADB]$

Subproof of 7. Suppose the conditional's antecedent. Then suppose that ${\sim}K_ADB$. Given Alice's observation of Betty, $K_A{\sim}DB$. From the conditonal's antecedent, K_AS, so $K_A(S \;\&\; {\sim}DB)$ by K1. Hence K_ADA by K1 and the definition of S. By the conditional's antecedent, ${\sim}K_ADA$. So, by reductio, $(K_AS \;\&\; {\sim}K_ADA) \rightarrow K_ADB$. Betty follows this inference of the conditional and so knows the conditional.

8. K_BK_ADB 6, 7, K1

9. K_BDB 8, K2, K1

10. $CB \leftrightarrow K_BDB$ Dispositions to wash

11. CB 9, 10

This argument shows how an act may arise from common knowledge. An act in a game may similarly arise from common knowledge. Later sections show how that happens. The remainder of this section further explains common knowledge.

Aumann (1976) analyzes common knowledge in terms of epistemic possibility. He defines communal possibility for a population in terms of epistemic possibility for members of the population. At a world ω, $M(\omega)$ stands for the set of communally possible worlds. A world ω' is communally possible if and only if within the population a path of epistemic accessibility goes from ω to ω'. More precisely, $\omega' \in M(\omega)$ if and only if for some n-tuple of members of the population (not necessarily n distinct members), at ω for the first member a world is possible such that at it for the second member a world is possible such that at it for the third member a world is possible … such that at it for the nth member ω' is possible. Aumann proves that in a world ω, a proposition p is common knowledge if and only if p is true in each element of $M(\omega)$.

Using communal possibility, one may conduct inferences about common knowledge without working through an infinite hierarchy of mutual knowledge.

Aumann's result linking communal possibility and common knowledge relies on idealizations. An agent may lack the concept of knowledge and so may not think about other agents' knowledge. Then truth in all communally possible worlds does not suffice for common knowledge. The move from communal possibility to common knowledge assumes that agents are ideal and are aware of other agents' epistemic possibilities.

To illustrate Aumann's theorem about common knowledge, I apply it to the example of the two travelers. Simplifying, there are four possible worlds distinguished according to whether Alice and Betty have clean or dirty faces. Figure 1 presents these worlds.

Worlds	1	2	3	4
A's face	*C*	*D*	*C*	*D*
B's face	*C*	*C*	*D*	*D*

Figure 1. Possible Worlds

One may flesh out these possible worlds by inferring from their features the relevant features of Alice's and Betty's knowledge. For instance, using background information the story provides, in world 1 one may infer that Alice knows that Betty's face is clean. The worlds include knowledge and not just facial states, although the worlds' representations mention just facial states.

Before the conductor's announcement, epistemic possibilities for the two women are: $P_A(4) = \{3, 4\}$, $P_B(4) = \{2, 4\}$. Communal possibilities are: $M(4) = \{1, 2, 3, 4\}$. In world 4, world 1 is communally possible because world 3 is possible for Alice, and, in world 3, world 1 is possible for Betty.

After the conductor's announcement, epistemic possibilities are: $P_A(4) = \{3, 4\}$, $P_B(4) = \{2, 4\}$. Communal possibilities are: $M(4) = \{2, 3, 4\}$. Possibilities do not change for Alice and Betty. They already knew someone's face is dirty. However, world 1 is no longer communally possible. In world 3, where only Betty's face is dirty, it is no longer possible for Betty that everyone's face is clean. In that world she sees that Alice's face is clean and infers that her own face is dirty. By Aumann's principle, it is common knowledge that someone's face is dirty because that proposition is true in every communally possible world.

After Alice does not use her turn for the lavatory, epistemic possibilities are: $P_A(4) = \{3, 4\}$, $P_B(4) = \{4\}$. Communal possibilities are: $M(4) = \{3, 4\}$. For Alice, it is still epistemically possible that her face is clean. However, Betty knows that she has a dirty face and so that both women do. Hence it is not communally possible that Betty's face is clean. That Betty's face is dirty is true in every world communally possible. Therefore by Aumann's principle it is common knowledge, not just that someone's face is dirty, but that Betty's face is dirty.

As noted, accounts of common knowledge involve ideal agents. To make the accounts more realistic, one may switch from knowledge to belief and then from belief to reasons for belief. Lewis (1969: Chap 2) shows how to do this during a treatment of common knowledge that grounds his theory of convention. He starts with an account of states of affairs that provide reasons for beliefs and are sources of common knowledge defined in terms of reasons. A state of affairs *A* that provides reasons for belief that a proposition is true indicates the truth of the proposition. Lewis defines indication as follows: "Let us say that *A indicates* to someone *x* that ____ if and only if, if *x* had reason to believe that *A* held, *x* would thereby have reason to believe that ____" (52–53).

A state of affairs that indicates a proposition's truth may generate common knowledge in a population that shares the same inductive standards and methods of reasoning. For example, a public announcement that *p* may indicate truth in a way that generates common knowledge. Lewis defines common knowledge this way: "Let us say that it is *common knowledge* in a population *P* that ____ if and only if some state of affairs *A* holds such that:

(1) Everyone in *P* has reason to believe that *A* holds.

(2) *A* indicates to everyone in *P* that everyone in *P* has reason to believe that *A* holds.

(3) *A* indicates to everyone in *P* that ____" (56).

According to this definition, common knowledge that *p* exists in a population when the members of the population have an infinite hierarchy of reasons for mutual beliefs. Some authors say that members of the population have implicit or tacit mutual beliefs because their beliefs do not ascend the whole hierarchy of reasons. Their beliefs ascend only a few steps up the hierarchy. Also, their beliefs may not amount to knowledge. Their beliefs may be false despite being justified by reasons in the hierarchy. However, in ideal cases their beliefs constitute knowledge, and their beliefs ascend the whole hierarchy of reasons.

Cubitt and Sugden (2003) reconstruct Lewis's account of common knowledge to show precisely how indication yields common knowledge. They show how an event that indicates a proposition, such as a public announcement that the proposition holds, generates reasons for beliefs that amount to common knowledge of the proposition in Lewis's sense.

First, Cubitt and Sugden restate Lewis's definition of common knowledge using their notation to express reasons and indication. $R_i(x)$ expresses that *i* has reason to believe that *x*. For a state *A*, $A ind_i x$ expresses that *A* indicates to *i* that *x*, that is, *A* gives *i* a reason to believe that *x*. Consequently, $A ind_i x$, if and only if in virtue of *A*'s holding $R_i(x)$. Cubitt and Sugden define common knowledge of a proposition *x* in an a population *P*, that is, $rP(x)$, using a hierarchy of reasons in place of a hierarchy of knowledge. According to their definition, $rP(x)$ if and only if for all persons $i, j, k, \ldots$ in *P*, $R_i(x)$, $R_i(R_j[x])$, $R_i(R_j[R_k(x)])$, ….

Next, they introduce the type of indication that generates common knowledge defined in terms of reasons. A state of affairs *A* is a *reflexive common indicator* that *x* in a population *P* if and only if:

(C1) For each person *i* in *P*, if *A* holds, then $R_i(A \text{ holds})$.

(C2) For all persons *i*, *j* in *P*, $A ind_i R_j(A \text{ holds})$.

(C3) For each person *i* in *P*, $A ind_i x$.

(C4) For all persons *i*, *j* in *P* and for each proposition *y*, if $A ind_i y$, then $R_i(A ind_j y)$.

Then they adopt some principles concerning indication. They use the following two principles to derive Lewis's claim about indication's giving rise to common knowledge.

(A1) For each person *i*, for all states of affairs *A*, and for each proposition *x*, if $[R_i(A \text{ holds}) \; \& \; A ind_i x]$, then $R_i(x)$.

(A6) For all persons *i*, *j*, for all states of affairs *A*, *A*', and for each proposition *x*, if $[A ind_i R_j(A' \text{ holds}) \; \& \; R_i(A' ind_j x)]$, then $A ind_i R_j(x)$.

Lastly, they show that *A*'s holding and being a reflexive common indicator in *P* that *x* yields common knowledge in *P* that *x*. This result establishes Lewis's claim about the rise of common knowledge from states of affairs that indicate truth.

Consider any state of affairs A, any proposition x, and any population P. Suppose that A holds and that in P the state A is a reflexive common indicator that x.

1.	R_i(A holds)	Assumption, C1
2.	$Aind_iR_j$(A holds)	Assumption, C2
3.	$Aind_ix$	Assumption, C3
4.	$R_i(x)$	1, 3, A1
5.	$R_i(Aind_jx)$	3, C4 ($y = x$)
6.	$Aind_iR_j(x)$	2, 5, A6 ($A = A'$)
7.	$R_i[R_j(x)]$	1, 6, A1 ($x = R_j(x)$)
8.	$R_i[Aind_jR_k(x)]$	6 ($j = k$), C4 ($y = R_k(x)$)
9.	$Aind_iR_j[R_k(x)]$	2, 8, A6 ($x = R_k(x)$, $A = A'$)
10.	$R_i[R_j(R_k[x])]$	1, 9, A1 ($x = R_j(R_k[x])$)
11.	$R_i[Aind_jR_k(R_l[x])]$	9 ($j = k$, $k = l$), C4 ($y = R_k(R_l[x])$)

and so on.

Lines 4, 7, 10, … establish the theorem because they amount to $rP(x)$, namely, common knowledge that x in the population P.[6]

This argument concerns the origin of common knowledge taken as a hierarchy of reasons. When the reasons warrant beliefs and agents are ideal, the hierarchy of reasons yields common knowledge taken as a hierarchy of mutual knowledge.

[6] To remove the ellipsis, one may use mathematical induction on the number of agents in a formula of the form of lines 4, 7, and 10. Let $F(n)$ be a formula of that form involving n agents. Establishing line 4 establishes $F(n)$ for n = 1. Then for any n, assume $F(n)$ and show $F(n + 1)$ using the method for going from line 4 to line 7 and from line 7

This section illustrates inferences involving the nature, origin, and consequences of common knowledge. Such inferences are complex. Formal epistemology analyzes them with precision and rigor. It may use semantic methods involving set theory, as does Aumann (1976), or syntactic methods involving modal logic, as does Halpern (2003).[7] Any rigorous and perspicuous analysis assists evaluation of reasoning that involves common knowledge.

3. Common Knowledge in Games of Strategy

Game theory treats agents' interactions. Games are situations in which the consequences for an agent of his act depend on other agents' acts. An agent's reasons for his acts include his beliefs about their consequences. In games, those beliefs depend on the agent's beliefs about the beliefs of other agents. To simplify, this paper assumes that the relevant beliefs amount to knowledge (with certainty). It examines an agent's knowledge about the knowledge of other agents. In particular, it examines common knowledge's effect on acts in noncooperative games of strategy. These are games in which the players act independently because they lack opportunities for joint action.

In ideal games, agents are cognitively ideal. Their reasons for beliefs yield common knowledge, and that knowledge justifies their participation in equilibria. For example, consider the sequential game in Figure 2. The players are *A* and *B*. Player *A* starts. Each player at a turn may choose Across or Down. Pairs of numbers stand for utilities of outcomes. The first number of a pair for an outcome is the outcome's utility for *A*, and the second number is the outcome's utility for *B*.

---

to line 10. That step completes the induction and establishes that $F(n)$ for all n. This generalization amounts to $rP(x)$.

```
A — Across  — B — Across  — (1, 1)
|             |
Down          Down
|             |
(2, 0)        (1, 2)
```

Figure 2. A Sequential Game

If the players reach *B*'s turn to move, *B* plays Down to gain 2 instead of 1. Player *A* foresees *B*'s move and plays Down to gain 2 instead of 1. The players' common knowledge of their game and their rationality generates *A*'s foresight. Player *A* knows that if she plays Across, *B* will play Down, because she knows that *B* knows the game and is rational. *A*'s knowledge of *B*'s knowledge of the game comes from their common knowledge of the game. That common knowledge may originate from a public reading of the game's rules.

Common knowledge's contribution to equilibrium in sequential games is well charted.[8] I investigate common knowledge's contribution to equilibrium in simultaneous-move games. These are games in which each player adopts a strategy without knowing other players' strategies. No player's strategy has a causal influence on another player's strategy.

In a game, a profile of strategies is a combination of strategies with exactly one strategy for each player. In ideal games, a Nash equilibrium is a profile of strategies such that each player's strategy maximizes utility given the profile. Idealizations make supposition of a profile amount to supposition of knowledge of the profile. Showing that players realize a Nash equilibrium requires showing that each player's participation is rational.

---

[7] Predominantly, economics takes the semantic approach, and computer science takes the syntactic approach.

The Stag Hunt, a game that Figure 3 presents, illustrates an effect of common knowledge in a simultaneous-move game. The profile (Up, Left) represents both players hunting stag, and the profile (Down, Right) represents both players hunting rabbit.

	Left	Right
Up	2, 2	0, 1
Down	1, 0	1, 1

Figure 3. The Stag Hunt

Both (U, L) and (D, R) are Nash equilibria. (U, L) is attractive because it is the efficient Nash equilibrium. However, each player participates in the profile only if confident that the other will.[9] Row is rational, but may wonder whether Column is rational and will do L. Column is rational, but may wonder whether Row knows that she is and so will do U in response to her doing L. Any missing link in the hierarchy of mutual knowledge of rationality creates a doubt about the wisdom of participating in (U, L). The players' common knowledge of their game and their rationality banishes doubts of this sort. If both players know that both hunt stag, then, for each, hunting stag maximizes utility. The profile (U, L) is the outcome.[10]

The classical treatment of simultaneous-move games, which I pursue, makes many idealizations. Von Neumann and Morgenstern (1944: 146–148), for instance, assume that

---

[8] In a sequential game, common knowledge supports a roll-back equilibrium of the game. It grounds the backwards induction that yields the roll-back equilibrium. Controversies concern the interpretation of behavior that deviates from the equilibrium.

[9] Harsanyi and Selten (1988) propose using risk-dominance to select a Nash equilibrium in such games. It sanctions the efficient Nash equilibrium only if participation maximizes expected utility.

players have knowledge of the theory of rationality and use it to figure out each other's strategies. Each player knows how the other players apply the theory of rationality to their game. Von Neumann and Morgenstern implicitly assume that agents have common knowledge of all relevant facts about their game. They do not explicitly state their common knowledge assumptions because when a text presents a game, a reader naturally assumes that the players have the same knowledge that the reader acquires as he learns about the game. As a background assumption, the reader supposes that all the players know what he knows about the game. Because the reader knows every relevant feature of the game, the background assumption entails that the players know every relevant feature of the game, too. What a player knows is a relevant feature of the game because it affects the player's behavior. The reader knows every relevant fact that any player knows. Hence, by assumption, the players all know every relevant fact that any player knows.

This assumption is enough to generate common knowledge of any relevant fact that any agent knows. That is, the agents in a game have common knowledge of a relevant proposition if at least one agent knows the proposition. Common knowledge arises from the Principle of Positive Introspection (the KK Principle) together with the assumption that each agent knows any relevant fact that another agent knows, including facts about agents' knowledge. To see this, suppose that an agent i knows the relevant fact that p, that is, K_ip. Then by assumption, every agent j knows that p, that is, K_jp. Because agent i satisfies the standard S5 principles of knowledge, including Positive Introspection, he knows that he knows that p. Hence every agent j knows that agent i knows that p. The same reasoning applies to K_jp. Thus, for all i, j, K_jK_ip. This result also holds for a proposition about an agent's knowledge of p. Consequently, for any

[10] Fagin et al. (1995: Chap. 6) argue that common knowledge is necessary for the efficient Nash equilibrium in a similar game of coordinated attack.

n-tuple of agents (not necessarily distinct), $K_1K_2 \ldots K_np$. The set of such strings entails the levels of mutual knowledge that constitute common knowledge that p.

To illustrate, I return to the example of Section 2. Consider Alice and Betty's common knowledge that S. A tree representing this common knowledge as levels of mutual knowledge begins as in Figure 4.

S

/ \

K_AS K_BS

/ \ / \

K_AK_AS K_BK_AS K_AK_BS K_BK_BS

Figure 4. A Tree Representing Common Knowledge

This tree continues level after level. Each level below S represents a level in the hierarchy of mutual knowledge that constitutes common knowledge that S. Instead of constructing the tree level by level, one may construct it path by path. Each path is infinitely long. The string K_AS, K_AK_AS, … represents the leftmost path of the tree, omitting S. Using K_AS and the generalization that for any p, $\exists iK_ip \rightarrow \forall j\forall iK_jK_ip$, one may derive for every path starting with K_AS a string representing it. Similarly, one may derive for every path starting with K_BS a string representing it. The collection of strings then represents the hierarchy of levels of mutual knowledge. The collection thus represents common knowledge that S. Cubitt and Sugden take advantage of this representation of common knowledge. They define common knowledge (in terms of reasons) using strings rather than levels.

In ideal games, players have common knowledge of their game and their rationality. It is as if all know that all have read the same description of the game and the players. They have the common knowledge that the description's public reading generates.

Rationality is a dispositional property of options. A rational option is one that if realized has certain desirable properties such as maximizing utility. A self-ratifying option has the dispositional property of maximizing utility if realized. Ratification calculates an option's utility using causal decision theory but under the assumption that the option is realized.

Von Neumann and Morgenstern assume that rational, informed players achieve an equilibrium. Their standard of rationality is utility maximization, but they apply it under the assumption that a strategy profile is realized. Because a player knows his own part of a strategy profile, his maximizing utility given a profile requires his strategy's self-ratification.

Bicchieri (2004) and others observe that players' common knowledge of their game's payoff matrix and their utility maximization is not sufficient for participation in a Nash equilibrium. It supports only a rationalizable strategy profile, that is, one remaining after iteratively eliminating strictly dominated strategies. This paper uses players' common knowledge of their game and their rationality to support the realization of a Nash equilibrium. Players' common knowledge may go beyond common knowledge of their payoff matrix. It may include knowledge of players' psychologies. One may also replace common knowledge that they maximize utility with common knowledge that they adopt self-ratifying strategies.

4. Support for a Nash Equilibrium

This section begins an argument that in some simultaneous-move games, it is rational for players to participate in a Nash equilibrium. It argues that ratification supports a Nash equilibrium given suitable common knowledge. Although the argument is not general, it illustrates principles that a general argument may employ.

One method of deriving an equilibrium's realization from players' reasoning assumes that their reasoning about strategies is bounded and takes place in stages. Cognitive limits force players to process bits of information sequentially instead of all at once. Harsanyi and Selten's (1988) tracing procedure and Skyrms's (1990) deliberational dynamics, for example, take this tack. However, their principles of bounded rationality do not extend to ideal cases. General principles of rationality cover ideal cases, too. They cover the strategic reasoning of ideal agents. I treat ideal games because in them strategic reasoning displays the structure of strategic reasons for acts.

The dynamic nature of decisions in games suggests that equilibrium rather than utility maximization directs rational agents. Should one shelf the principle to maximize utility and substitute the principle to do one's part in a unique Nash equilibrium? A single set of decision principles good for all decision problems gives a theory of rationality explanatory power. To support a Nash equilibrium, it is best to show that independently motivated principles of rationality support participation in a Nash equilibrium, that is, adoption of a Nash strategy. Showing that a Nash equilibrium follows from players' adopting self-ratifying strategies unifies decision theory and game theory.[11]

Deriving realization of a Nash equilibrium from players' strategic reasoning confronts the interdependence of the probabilities players assign to each other's strategies. To apply the principle to maximize expected utility, a player needs the expected utility of each of his

---

[11] In the games treated, I assume that ratification and Nash equilibrium are adequate accounts of rationality and equilibrium. Weirich (1998) advances more general accounts of rationality and equilibrium.

strategies. To obtain a strategy's expected utility, he needs the probabilities of other players' strategies. For every player, the probability of a strategy depends on the probabilities of the other players' strategies. So the probabilities of players' strategies are interdependent. The principle of ratification overcomes this problem. It assumes that an agent adopts a strategy and then uses the probability of another agent's strategy under that assumption. The assumption gives probability assignments a foothold.[12]

Aumann and Brandenburger (1995) observe that utility maximization and knowledge of the profile realized support realization of a Nash equilibrium. If all players know the profile realized and maximize utility, then they achieve a Nash equilibrium. Only a Nash equilibrium is such that, given knowledge of its realization, each strategy in it maximizes utility. Aumann and Brandenburger support the realization of some Nash equilibrium or other, and not the realization of a particular Nash equilibrium. If a game has multiple Nash equilibria, an explanation of the players' realization of a particular equilibrium must show how the players coordinate to realize that equilibrium. That they realize a Nash equilibrium does not explain the realization of the particular Nash equilibrium they realize.

Aumann and Brandenburger's result does not depend on common knowledge. However, as they acknowledge, their result does not explain realization of a Nash equilibrium. Suppose that utility-maximizing players have knowledge of the profile realized and realize a Nash equilibrium. Their knowledge does not by itself explain their realization of a Nash equilibrium. Suppose that if the players were to realize a non-equilibrium profile, they would have thought that they were making best responses to others' strategies. They would have had mistaken beliefs about the strategies of other players. Their beliefs about other players' strategies would

---

[12] Myerson (1991: 4, 114) describes the circle of strategic reasoning that arises when applying expected utility maximization in games of strategy. Weirich (2004: Chap. 9) shows that ratification breaks the circle of strategic

not have responded to the hypothetical change in the profile realized. Then their knowledge of the profile realized does not explain their realization of a Nash equilibrium. An explanation of a Nash equilibrium's realization requires that players have robust abilities to know other players' strategies.

To explain players' realization of a Nash equilibrium in an ideal game, I assume that they know the profile realized whatever it is. That is, for every profile, each player knows that the profile is realized if it is realized. This hypothetical knowledge covers every profile and not just the profiled realized. I also assume that each player maximizes utility given the profile realized whatever it is. Consequently, players adopt self-ratifying strategies.

Knowledge of the profile realized, whatever it is, comes from a player's knowledge of his own strategy, whatever it is, and knowledge of the other players' response to his strategy. A player's knowledge of responses to his strategies I call prescience. Prescience explains knowledge of the profile realized and also knowledge of any profile hypothetically realized. Prescience is a product of players' common knowledge of their game and their rationality. Their common knowledge explains the knowledge of conditionals that constitutes prescience. Common knowledge explains robust knowledge of the profile realized and thereby contributes to an explanation of a Nash equilibrium's realization.

This section shows that common knowledge supports prescience and that prescience, together with ratification, supports a Nash equilibrium in an ideal game with a unique Nash equilibrium. It puts aside coordination to achieve a particular Nash equilibrium when several exist.

The account of an agent's strategic reasoning starts with knowledge about the game and the

---

reasoning.

players. So that an agent’s deliberations explain his choice, they do not begin with direct knowledge of his choice and direct knowledge that his choice is rational. An agent may have indirect foreknowledge of his choice. He may obtain that foreknowledge in the course of deliberations. Each agent knows that the other agents are certain that he will make a best response to them. This assumption does not give an agent direct knowledge of his choice. It gives him knowledge of other agents' choices conditional on knowledge of his choice. An agent knows that others are certain that he will make a best reply, and that they are rational, ideal agents, and so justified in their certainty. Initially, he considers their certainty to be justified but nonetheless possibly mistaken. He does not know directly that others know that he will make a best reply, nor does he know directly that he will make a best reply.

In a simultaneous-move game, no player observes the others' moves before he moves. A player may infer the other players' moves, however. A player's choice and other players' choices arise from the theory of rationality. Given a player's choice, he infers the choices of other players. He knows that other players respond rationally to what he does. The assumption that a player has knowledge of his own choice and a rational response to it grounds his inferences about other players' choices.

Agents who realize a game's unique Nash equilibrium coordinate strategies to realize that profile although they do not have the common goal of realizing the profile. They each realize their parts because each has the goal of maximizing utility given background assumptions. That is enough for coordination because each responds to evidence about the others' acts. Each acts because of expectations about the others. Realization of the Nash equilibrium is epistemic coordination on a strategy profile. The players respond inferentially to a common set of circumstances: the game and their knowledge of it and each other.

The argument for participation in a Nash equilibrium uses common knowledge of the game and the rationality of players. To begin, it clarifies knowledge of the game and knowledge of rationality. Knowledge of the game includes knowledge of the payoff matrix but may include more. It may include relevant facts about the players such as their knowledge and their responses to out-of-equilibrium strategies. Knowledge of the players' rationality includes knowledge that they conform to principles of rationality. Utility maximization is a common principle of rationality, but I substitute the principle of ratification as a refinement. According to it, players adopt a strategy that maximizes utility on the assumption that it is adopted. This principle assists reasoning that supports participation in a Nash equilibrium.

As an example, this section treats Matching Pennies, a two-person game in which each player has two strategies. Figure 5 presents the game's payoff matrix.

	Left	Right
Up	2, 0	0, 2
Down	0, 2	2, 0

Figure 5. Matching Pennies

The game has only a mixed-strategy Nash equilibrium (1/2, 1/2). The players realize this equilibrium if each flips a coin to select a pure strategy from the payoff matrix. What reasoning supports this equilibrium? The players have common knowledge of conditionals concerning their strategies such as, "If Up, then Right." In consequence, Row's mixed strategy 1/2 is his only ratifiable strategy. This is common knowledge. Given that each agent adopts a ratifiable strategy, each participates in the equilibrium. This is common knowledge because it follows from premisses that are common knowledge.

Common knowledge supports a player's participation in an equilibrium by supporting knowledge of his opponent's strategy. The opponent's strategy is supported by her knowledge of the first player's strategy. Each knows the other's strategy by an inference. The inference involves common knowledge. Common knowledge that just one profile has self-ratifying strategies gives each agent a reason to do his part in the profile.

A pure strategy has the same expected payoff as a player's Nash strategy given that his opponent participates in the equilibrium. Why does the player adopt his Nash strategy? The reason appeals to out-of-equilibrium behavior, which prescience covers. Adopting a pure strategy gives an agent evidence that his opponent adopts a best response and so a deviation from her equilibrium strategy. Given her deviation, his pure strategy does not maximize utility.

Row's knowledge of Column's response to a deviation involves common knowledge. Knowledge of conditionals such as, "If Up, then Right," arises from common knowledge. A representation of the players' reasoning gains depth by explaining the reasoning that generates that knowledge. It comes from an agent's knowledge of his own strategy and an agent's adopting a best response to an opponent's strategy, and all this being common knowledge. If Row adopts Up, then he knows this. If he knows his strategy, then Column knows that he knows. If Column knows that, then Column knows his strategy. Hence Column adopts a best response. Column knows her response. Hence Row knows her response. Everything relevant that anyone knows is common knowledge. Hence it is common knowledge that if Row adopts Up, then Column adopts Right. By such reasoning, each player attains prescience of responses. Each knows for each strategy the response of the other. Common knowledge explains prescience and so realization of a Nash equilibrium. The next section elaborates this support for a game's unique Nash equilibrium.[13]

5. Epistemic Logic and Nash Equilibrium

This section illustrates derivation of a Nash equilibrium's realization from agents' common knowledge of their game. Support for participation in a Nash equilibrium entertains counterfactual worlds, which the antecedents of counterfactual conditionals introduce. Analysis of counterfactual conditionals, both indicative (evidential) and subjunctive (causal), introduces relations between worlds in addition to the accessibility relation representing epistemic possibility. To represent strategic reasoning, one may use a modal system representing knowledge combined with a logic of counterfactual conditionals involving two distance relations between possible worlds, one relation for indicative conditionals and another relation for subjunctive conditionals.[14]

A conditional having a counterfactual antecedent is true if its consequent holds in a world minimally revised to accommodate the antecedent.[15] In ideal games, strategic reasoning involving counterfactual conditionals attributes rationality to agents. Minimal revision preserves insofar as possible assumptions about agents' rationality. The nearness-conditionals representing agent's responses to strategies in an ideal simultaneous-move game preserve, first, the predictive power of agents and, second, the rationality of agents. I use the corner > as a connective forming

[13] A version of the section’s argument is in Weirich (Forthcoming).
[14] Supposition of a conditional’s antecedent need not add a proposition to one's knowledge. It may entertain the proposition without the support required for knowledge, unless background idealizations make the supposition carry knowledge of its realization. Also, knowledge of the conditional (if *p* then *q*) need not carry knowledge of *q* in a world where *p* is true. Knowledge of the conditional entails that *p*'s realization brings *q*'s realization, but knowledge of *q*'s realization is another matter.

counterfactual conditionals. Context distinguishes the corner from the symbol for the greater-than relation. Sentences flank the corner, whereas names and variables flank the symbol for the greater-than relation.

In an ideal game all relevant knowledge is common knowledge. A single knowledge operator *K* may represent each agent's knowledge and also the agents' common knowledge.[16] The knowledge operator *K* has its standard interpretation in terms of a set of worlds and an accessibility relation for them. The semantics for the corner > use a selection function that given a world and proposition yields the nearest world in which the proposition is true. A corner-conditional is true at a world *w* just in case the consequent is true in the world selected given *w* and the antecedent. For simplicity, this section uses a single selection function underwriting only indicative conditionals.

The argument concludes that in an ideal version of Matching Pennies, where players independently pick strategies, they realize the Nash equilibrium in mixed strategies. The payoff matrix in Figure 5 depicts the game. There are two players, Row and Column. Row's pure strategies are Up and Down. Column's pure strategies are Left and Right. A mixed strategy for Row, specifies a value for *p*, the probability of Up, using an integer between 0 and 100 taken as a percentage. Similarly, a mixed strategy for Column specifies a value for *q*, the probability of Left. Each player realizes exactly one mixed strategy.

The argument includes premisses concerning players' utilities and choices. These premisses use *u* as a conditional utility function. The conditional utility $u(o, s)$ is the utility of an option *o* given a state *s*. The utility function *u* uses names of propositions and variables as place-holders for names of propositions. In the context of the function *u*, a symbol for a proposition

[15] I assume that only one world is a minimal revision of another.

serves as a name of a proposition. The context converts a sentence expressing a proposition into a standard name of the proposition, for instance, a that-clause name of the proposition. In a game an option's utility depends on states concerning other agents' strategies.

The premisses are principles of rationality and principles specifying nearest worlds. First is a general principle of conditional utility, simplified for games where players select strategies independently.[17]

Expected Utility: $u(o, o') = \sum_i p(s_i / o')u(o, s_i)$, where $\{s_i\}$ is a partition of states

A relevant state is an opponent's strategy. If an opponent's strategy s is known, that knowledge swamps the evidence o' provides about his strategy. Suppose that knowledge of a proposition makes the probability of the proposition equal to 1 given any supposition compatible with the proposition. Imagine that $K(p = 50\%)$. Then consider $u_C(q = 0\%, q = 0\%)$. By Expected Utility, it equals $p(p = 50\% / q = 0\%)u_C(q = 0\%, p = 50\%) + p(p \neq 50\% / q = 0\%)u_C(q = 0\%, p \neq 50\%) = u_C(q = 0\%, p = 50\%)$. Because the condition $p = 50\%$ is known, it replaces the condition $q = 0\%$. Another premiss generalizes this effect of knowing an opponent's strategy s.

Knowledge: $Ks > u(o, o') = u(o, s)$

Third is a decision principle.

[16] This representation conceals differences in the agents' acquisition of knowledge, however. One agent may infer that another agent knows that p. The other agent may use introspection to learn that she knows that p.

[17] In a simultaneous-move game, where an agent's options do not causally influence options of other players, causal decision theory simplifies the probabilities of states on which an option's utility depends. Let '$>_c$' be the connective for causal conditionals. Causal decision theory in the general case uses $P((o >_c s)/o')$ to weight the utility of an

Ratification: $o > \forall o'\ u(o, o) \geq u(o', o)$

An option o satisfying the consequent is self-ratifying. The principle of ratification applies when a self-ratifying option exists. It says that only a self-ratifying option is realized. Using knowledge of an opponent's strategy yields a simplification of Ratification: $Ks > (o > \forall o'\ u(o, s) \geq u(o', s))$.

Fourth is a principle of self-knowledge.

Awareness: $o > Ko$

In an ideal game, an agent's knowledge that he realizes an option yields common knowledge of the option's realization. A deeper-going justification of Nash equilibrium than this section provides, derives the common knowledge Awareness expresses from common knowledge of other sorts concerning the game and the players.

The argument uses knowledge and conditionals taken with respect to the world in which the game occurs. To simplify, it omits indexing the knowledge operator K to that world and also omits indexing the corner > to that world. In counterfactual conditionals, the index for an embedded occurrence of K and an embedded occurrence of > is settled implicitly by the semantics of the conditionals.

The argument also has some premisses about conditionals that express the effects on conditionals of distance between worlds. One premiss asserts that given an opponent's strategy,

---

option o given that a state s obtains if an option o' is realized. In a simultaneous-move game, $P((o >_c s)/o') = P(s/o')$ because o has no causal influence on s. The simplified formula for the relevant probability omits option o.

an agent adopts a best response. This conditional holds no matter how deeply embedded it is in other conditionals.

> Best Response: In all contexts Column adopts a best response to Row: $p = x > (q = y > \forall z\ u_C(q = y, p = x) \geq u_C(q = z, p = x))$. Likewise, in all contexts Row adopts a best response to Column: $q = y > (p = z > \forall w\ u_R(p = z, q = y) \geq u_R(p = w, q = y))$.

When the conditional about Row's best response is embedded in a supposition about Row's strategy, the result is this: $p = x > (q = y > (p = z > \forall w\ u_R(p = z, q = y) \geq u_R(p = w, q = y))$. To illustrate, suppose that $p = 50\%$. Then make the further supposition that $q = 100\%$. Because $p = 50\%$ is not a best response to $q = 100\%$, adding the supposition that $q = 100\%$ overturns the initial supposition that $p = 50\%$. Although $q = 100\%$ is a best response to $p = 50\%$, it is not a best response to the value of p it indicates, namely, $p = 100\%$. By Best Response, $p = 50\% > (q = 100\% > p = 100\%)$.

A related premiss retains a supposition about an agent's strategy if it is a best response to a further supposition about the opponent's strategy.

> Nearness: For Row, $p = x > (q = y > (\forall z\ u_R(p = x, q = y) \geq u_R(p = z, q = y) > p = x))$. An analogous generalization holds for Column.

To illustrate for Row, suppose that $p = 50\%$. Because that strategy is a best response to $q = 50\%$, supposition that $q = 50\%$ retains the original supposition. By Nearness, $p = 50\% > (q = 50\% > p = 50\%)$.

The following proof that agents realize the mixed-strategy Nash equilibrium displays the principal inferential steps but compresses other steps. Its main step shows that only the agents' Nash strategies are self-ratifying. Hence following Ratification, the agents realize those strategies.

First, the proof establishes that each player's Nash strategy is self-ratifying. It treats Row's Nash strategy step by step. Similar steps apply to Column's Nash strategy. To show that Row's Nash strategy $p = 50\%$ is self-ratifying, one must show that $\forall x\ u_R(p = 50\%, p = 50\%) \geq u_R(p = x, p = 50\%)$. Consider, for example, the alternative strategy $p = 100\%$. Its conditional utility $u_R(p = 100\%, p = 50\%)$ depends on what the condition $p = 50\%$ shows about Column's strategy. Given $p = 50\%$, $K(p = 50\%)$ by Awareness. Hence all Column's strategies have the same expected utility. Also, $p = 50\% > (q = 50\% > p = 50\%)$ by Nearness. So $q = 50\%$ is self-ratifying. Moreover, only $q = 50\%$ is self-ratifying. Take the alternative $q = 100\%$. By Best Response, $p = 50\% > (q = 100\% > p = 100\%)$. Column's strategy $q = 100\%$ is an inferior response to $p = 100\%$. So it is not self-ratifying. In general, alternatives to $q = 50\%$ are not self-ratifying. Therefore by Ratification, $q = 50\%$. Thus $K(q = 50\%)$ by Awareness. So all Row's strategies have the same expected utility. Hence $u_R(p = 50\%, p = 50\%) = u_R(p = 100\%, p = 50\%)$. Similarly, the equality holds when any other alternative strategy replaces $p = 100\%$. Therefore $p = 50\%$ is self-ratifying.

Next, the proof establishes that nonNash strategies are not self-ratifying. Consider Row's options. They reduce to three: $p > 50\%$, $p < 50\%$, and $p = 50\%$. Row realizes the coarse-grained option $p > 50\%$ just in case he realizes a fine-grained option satisfying the inequality, and similarly for $p < 50\%$. Suppose that Row realizes the option that $p > 50\%$. Then $K(p > 50\%)$ by Awareness. Because $U_C(q = 0\%, p > 50\%) > U_C(q > 0\%, p > 50\%)$, Knowledge implies that

$U_C(q = 0\%, q = 0\%) > U_C(q > 0\%, q = 0\%)$. Hence, by Ratification, $q = 0\%$, as all other options are not self-ratifying. So by Awareness, $K(q = 0\%)$. Because $U_R(p > 50\%, q = 0\%) < U_R(p = 0\%, q = 0\%)$, Knowledge implies that $U_R(p > 50\%, p > 50\%) < U_R(p = 0\%, p > 50\%)$. So by Ratification, it is not the case that $p > 50\%$. Similarly, the supposition that $p < 50\%$ leads to a contradiction. So by Ratification, $p = 50\%$. The same steps establish that $q = 50\%$. Realization of the two Nash strategies constitutes realization of the game's Nash equilibrium.

Showing that players in an ideal version of Matching Pennies achieve the game's unique Nash equilibrium is just step toward a general account of realization of a Nash equilibrium in ideal simultaneous-move games. However, this simple example illustrates the role of common knowledge in generating prescience, which together with the principle of ratification yields a player's adoption of a Nash strategy and the players' realization of a Nash equilibrium.

References

Aumann, Robert. 1976. "Agreeing to Disagree." *Annals of Statistics* 4: 1236–1239.

Aumann, Robert and Adam Brandenburger. 1995. “Epistemic Conditions for Nash Equilibrium.” *Econometrica* 63: 1161–1180.

Bergmann, Michael. 2006. *Justification without Awareness*. Oxford: Oxford University Press.

Bicchieri, Cristina. 2004. "Rationality and Game Theory." In Alfred Mele and Piers Rawling, eds., *The Oxford Handbook of Rationality*, pp. 182–205. New York: Oxford University Press.

Binmore, Ken and Adam Brandenburger. 1990. "Common Knowledge and Game Theory." In Ken Binmore, *Essays on the Foundations of Game Theory*, pp. 105–150. Oxford: Blackwell.

Cubitt, Robin and Robert Sugden. 2003. "Common Knowledge, Salience and Convention: A Reconstruction of David Lewis' Game Theory." *Economics and Philosophy* 19: 175–210.

Fagin, Ronald, Joseph Halpern, Yoram Moses, and Moshe Vardi. 1995. *Reasoning about Knowledge*. Cambridge, MA: MIT Press.

Halpern, Joseph. 2003. *Reasoning about Uncertainty*. Cambridge, MA: MIT Press.

Harsanyi, John and Reinhard Selten. 1988. *A General Theory of Equilibrium Selection in Games*. Cambridge, MA: MIT Press.

Lewis, David. 1969. *Convention: A Philosophical Study*. Cambridge, MA: Harvard University Press.

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

Skyrms, Brian. 1990. *The Dynamics of Rational Deliberation*. Cambridge, MA: Harvard University Press.

von Neumann, John and Oskar Morgenstern. 1944. *Theory of Games and Economic Behavior*. Princeton, NJ: Princeton University Press.

Weirich, Paul. 1998. *Equilibrium and Rationality: Game Theory Revised by Decision Rules*. Cambridge: Cambridge University Press.

Weirich, Paul. 2004. *Realistic Decision Theory: Rules for Nonideal Agents in Nonideal Circumstances*. New York: Oxford University Press.

Weirich, Paul. Forthcoming. *Collective Rationality: Equilibrium in Cooperative Games*. New York: Oxford University Press.