Outguessing and Deception in Novel Strategic Situations

Vincent P. Crawford, University of California, San Diego

SESS Distinguished Lecture, Singapore Management University
9 November 2004

Overview

Many strategic situations in business, international relations, politics, or war are well approximated by what I will call *outguessing games*—two- or more-person situations of pure conflict in which some players want to match other players' actions, and other players want to avoid matching them

Game theory has a standard model of how people decide what to do in outguessing and other kinds of games:

> *Nash equilibrium* (often shortened to *equilibrium*) in which each player chooses an action that is best for himself, given correct expectations about other players' actions

Equilibrium makes clear predictions for outguessing games, which are often accurate when players have learned to predict others' responses from experience with analogous games; Walker and Wooders, "Minimax Play at Wimbledon," *AER* '01

But in novel strategic situations there may be no analogous games, and equilibrium must come from sophisticated strategic thinking rather than learning from direct experience

This makes equilibrium a less plausible assumption, and equilibrium predictions are often less reliable for initial responses to games than when learning is possible

This lecture describes a non-equilibrium model of initial responses to games that has emerged from some recent experimental work, and compares it with equilibrium as a model of behavior in outguessing games

The lecture begins with some simple—but not completely unrealistic—examples that illustrate the key strategic issues

It then compares equilibrium predictions in the examples with history, experimental data, or intuitions about strategic behavior, highlighting puzzles that equilibrium gets wrong

Next, it introduces a non-equilibrium model of initial responses based on "level-*k*" thinking, which is closer to strategic intuition and experimental evidence

In some games the level-*k* model's predictions coincide with equilibrium, so equilibrium predictions rest on weaker, more plausible assumptions; and are correspondingly more reliable

In other games the level-*k* model's predictions may deviate systematically from equilibrium, but in predictable ways

The lecture concludes by showing that in outguessing games, the level-*k* model deviates systematically from equilibrium

These deviations bring its predictions closer to evidence and intuition, resolving some of the puzzles left open by equilibrium analysis

The simplest outguessing game: Matching Pennies

In Matching Pennies, two players, Row and Column, choose simultaneously between two actions, Heads and Tails; Column wins if they match and Row wins if they mismatch

Assume players choose their actions with the goal of winning, which yields a reward (*payoff*) of 1, while losing yields -1

		Column	
		Heads	Tails
Row	Heads	1 -1	-1 1
	Tails	-1 1	1 -1

Matching Pennies

In an equilibrium, each player chooses his best action, given correct expectations about the other's action

If players choose only between Heads and Tails, Matching Pennies has no equilibrium, because any combination of actions includes one that is not best for one of them

But in Matching Pennies it is important to be unpredictable, so it is natural to interpret "choice" to include randomized (*mixed*) actions as well as unrandomized (*pure*) actions

Think of a mixed action as choosing the *probabilities* of Heads and Tails, and of a player's goal as maximizing his expected payoff, or here, the probability of winning

Mixed actions are less weird than they may seem because a player's action need only be unpredictable to others, not himself: his choice could be nonrandom, based on private discussions with subordinates, and his mixed action may represent others' uncertain expectations about his action

		Column	
		Heads (*q*)	**Tails**
Row	**Heads (*p*)**	1 -1	-1 1
	Tails	-1 1	1 -1

Matching Pennies

With mixed actions Matching Pennies has an equilibrium, in which each player plays each action with probability 1/2: $1p$ $-1(1-p) = -1p + 1(1-p)$ and $-1q +1(1-q) = 1q -1(1-q)$ (odd as p (q) is determined to make Column (Row) indifferent)

If players choose their actions with probabilities $p = q = 1/2$, and a player correctly anticipates his opponent's probability, then Heads and Tails yield him the same expected payoff, no pure or mixed action has higher expected payoff, and randomizing 50–50 (p or $q = 1/2$) is one of his best choices; this is a kind of "rational expectations" equilibrium, in which players form correct expectations not about a market aggregate but about each other's action distribution

$p = q = 1/2$ is the only equilibrium in Matching Pennies: If a player could predict a choice probability different than 1/2 for the other (say by observing "tells"), then one of his pure actions would yield a higher expected payoff; but we know that Matching Pennies has no equilibrium in pure actions

More interesting examples of outguessing games

In Matching Pennies equilibrium is a reasonable prediction even for initial responses, because the only issue the game raises is unpredictability, the game is symmetric, and the need to randomize 50-50 is easily grasped even by children

More complex outguessing games raise more subtle (and more interesting) strategic issues, and equilibrium becomes correspondingly less reliable in predicting initial responses

These include games with:

● More than two actions, as in Rock-Paper-Scissors

● More than two players, as in Keynes' famous "beauty contest" example (*The General Theory*, ch. 12), likening professional investment:

> . . . to those newspaper competitions in which the competitors have to pick out the six prettiest faces from a hundred photographs, the prize being awarded to the competitor whose choice most nearly corresponds to the average preferences of the competitors as a whole; so that each competitor has to pick, not those faces which he himself finds prettiest, but those which he thinks likeliest to catch the fancy of the other competitors, all of whom are looking at the problem from the same point of view.

● Differences across actions in social context, as in Keynes' beauty contest—in this case people's ideas of prettiness

● More than two actions, with differences across actions in the cultural or geographic "landscape", as in the popular modern game: Where's @$!#?&? [your "favorite" terrorist]

● More than two actions, with differences across actions in order and labeling as in Rubinstein, Tversky, and Heller's '93, '96 experimental Hide and Seek games:

> You and another student are playing the following game: Your opponent has hidden a prize in one of four boxes arranged in a row. The boxes are marked as follows: A, B, A, A. Your goal is, of course, to find the prize. His goal is that you will not find it. You are allowed to open only one box. Which box are you going to open?

In this game the framing (order and labeling) of the four locations is a tractable abstract model of a cultural or geographic landscape like those that play important roles in real Hide and Seek games, such as Where's @$!#?&?

● Differences across actions in payoffs, as in D-Day:

		Germans	
		Defend Calais	**Defend Normandy**
Allies	**Attack Calais**	1 -1	-2 2
	Attack Normandy	-1 1	1 -1

D-Day

In D-Day the payoffs have been "stretched" from realistic values to clarify the relationship to Matching Pennies:

- Attacking an undefended Calais (closer to England) is better for the Allies than attacking an undefended Normandy, and so better for the Allies "on average"

- Defending an unattacked Normandy is worse for the Germans than defending an unattacked Calais, and so worse for the Germans on average

- Differences across actions in payoffs, as in D-Day's ancient Chinese antecedent, Huarongdao, in which Cao Cao chooses between two roads, trying to avoid capture by Kongming (thanks to Duozhe Li for the reference to Luo Guanzhong's historical novel, *Three Kingdoms*):

		Kongming	
		Main Road	**Huarong**
Cao Cao	**Main Road**	3 -1	0 1
	Huarong	1 0	2 -2

Huarongdao

Here the payoffs have not been stretched; they assume:

- Cao Cao loses 2 and Kongming gains 2 if Cao Cao is captured

- Both Cao Cao and Kongming gain 1 by taking the Main Road (easier), whether or not Cao Cao is captured

Despite the different payoffs, D-Day's and Huarongdao's strategic structures are very close:

- Column (Row) player wants to match (mismatch)

- Main Road is better for both Cao Cao and Kongming on average, just as Attack/Defend Calais was for Allies/Germans

- There are no pure equilibria and there is a unique mixed equilibrium

● Huarongdao and D-Day are even more interesting if we add an opportunity to send a message about intentions before the actions are chosen, as in Kongming's fires along the road at Huarong and D-Day's Operation Fortitude:

An Inflatable "Tank" from Operation Fortitude

(Compare Nathan Rothschild's—probably apocryphal—pretense of having received early news of a British defeat at Waterloo, so that he could profit by buying British government securities at temporarily depressed prices)

Equilibrium versus history, data, or intuition

● With a payoff of 1 for winning, RTH's Hide and Seek game translates into:

		Seeker			
		A	B	A	A
Hider	A	1 0	0 1	1 0	0 1
	B	0 1	1 0	1 0	0 1
	A	0 1	0 1	1 0	0 1
	A	0 1	0 1	0 1	1 0

Hide and Seek

Record your intuitions about how to play as Hider or Seeker

Like Matching Pennies, Hide and Seek has a unique, mixed equilibrium, with equal probabilities on all four locations for both players

But RTH's ABAA framing of the locations is non-neutral in two ways: the *B* location is distinguished by its label and the two *end A* locations are inherently focal; together these two focalities distinguish *central A* as "the least salient location"

Equilibrium leaves no room for the non-neutral ABAA framing to influence people's choices

But in RTH's experiments, *central A* was most prevalent for both Hiders (37%) and Seekers (46%), even more prevalent for Seekers; as a result Seekers find a Treasure more than 25% of the time and have higher payoffs than in equilibrium

Puzzles unresolved by equilibrium:

- If Seekers are as smart as Hiders on average, why don't Hiders who are tempted to hide in *central A* realize that Seekers will be just as tempted to look there?

- Why do Hiders choose actions that, on average, allow Seekers to find them more than 25% of the time, when they could hold it down to 25% via the equilibrium mixed action (or even lower by hiding anywhere but *central A*)?

- D-Day

		Germans	
		Defend Calais (*q*)	**Defend Normandy**
Allies	**Attack Calais (*p*)**	1 -1	-2 2
	Attack Normandy	-1 1	1 -1

D-Day

Compare D-Day with Matching Pennies and record your intuitions about how to play as Allies or Germans (setting *p*, *q* = 0, 1, or if you prefer, somewhere in between)

The equilibrium p and q solve:

- $1p - 1(1-p) = -2p + 1(1-p)$, which yields $p = 2/5$
- $-1q + 2(1-q) = 1q - 1(1-q))$, which yields $q = 3/5$

This probably matches your intuition (at least qualitatively) for the Germans because their better-on-average action, Defend Calais, has probability $q > 1/2$

But it probably goes against your qualitative intuition for the Allies because their better-on-average action, Attack Calais, has probability $p < 1/2$

The equilibrium must be counterintuitive here because if the Allies tried to exploit the ease of attacking Calais in the obvious way (setting $p = 1$), and this was predictable, then the Germans could neutralize the exploitation by defending Calais for certain (setting $q = 1$), yielding the Allies -1

With the predictability that equilibrium assumes, Allies can exploit the ease of attacking Calais only by setting $p < 1/2$

The equilibrium p, 2/5, yields Allies payoff 1/5, greater than their equilibrium payoff of 0 in Matching Pennies

This principle seems too subtle to be identified in bridge textbooks or informal writing on strategy (but see vN-M '53; vN '53; Crawford and Smallwood, *Theory and Decision* '84)

Puzzle: The Allies' decision to attack Normandy and the Germans' decision to defend Calais were both (apparently) nonrandom; but equilibrium explains only in an unhelpful way, by giving the realized outcome probability 9/25

● D-Day (or Haurongdao) with a message about intentions

In D-Day (or Haurongdao) with a costless message about intentions from the Allies (or Kongming) to the Germans (or Cao Cao), all equilibria have the "sender" sending an uninformative message, which the "receiver" ignores

Otherwise the receiver would benefit by responding to the message; but such a response would hurt the sender, who would thus do better to make his message uninformative

Given this, equilibrium with a message reduces to the mixed equilibrium of the game without a message

But attempts to deceive opponents about one's intentions are ubiquitous in outguessing games; for example:

- The Allies faked preparations for an invasion at Calais

- Kongming had fires lit along the road at Huarong

Further, in both D-Day and Huarongdao:

- The sender correctly anticipated which message would fool the receiver and chose it nonrandomly, the deception succeeded, but the sender won in the less beneficial of the two possible ways

- The sender's message and subsequent action were part of a single, integrated strategy; and his action differed from the action he would have chosen with no opportunity to send a deceptive message

(An unimportant difference is that in D-Day the message was literally deceptive but the Germans "believed" it, either because they were credulous or, more likely, because they inverted it one too many times; while in Huarongdao the message was literally truthful, but Cao Cao inverted it

Three Kingdoms gives Kongming's rationale for sending a deceptively truthful message ("Have you forgotten the tactic of 'letting weak points look weak and strong points look strong'?") and Cao Cao's rationale for inverting it ("Don't you know what the military texts say? 'A show of force is best where you are weak. Where strong, feign weakness.'")

Cao Cao must have bought a used, out-of-date edition!

Puzzles unresolved by equilibrium:

- Why did the receiver allow himself to be fooled by a costless (hence easily faked) message from an *enemy*?

- Was it a coincidence that, in both Huarongdao and D-Day, the sender sent a message that fooled the receiver in a way that allowed him to win in the *less* beneficial of the two possible ways to win?

- And if the sender expected his message to fool the receiver, why didn't he reverse the message and fool the receiver in a way that allowed him to win in the *more* beneficial way? (Why didn't the Allies feint at Normandy and attack Calais? Why didn't Kongming light fires on the Main Road and ambush Cao Cao there?)

Resolving the puzzles with a non-equilibrium model of initial responses based on "level-*k*" thinking

I now describe a non-equilibrium model of initial responses that is closer to intuition and predicts initial responses better than equilibrium in a wide range of game experiments, and which helps to resolve the puzzles left open by equilibrium

Consider subjects' initial responses in Nagel's '95 *AER* "guessing games" (inspired by Keynes' beauty contest):
● 15-18 subjects simultaneously guess between [0,100]
● The subject whose guess is closest to a *p* (= 1/2 or 2/3), times the group average guess wins a prize, say $50
● The structure is publicly announced

Record your intuition about what to guess if *p* = 1/2, or 1/3

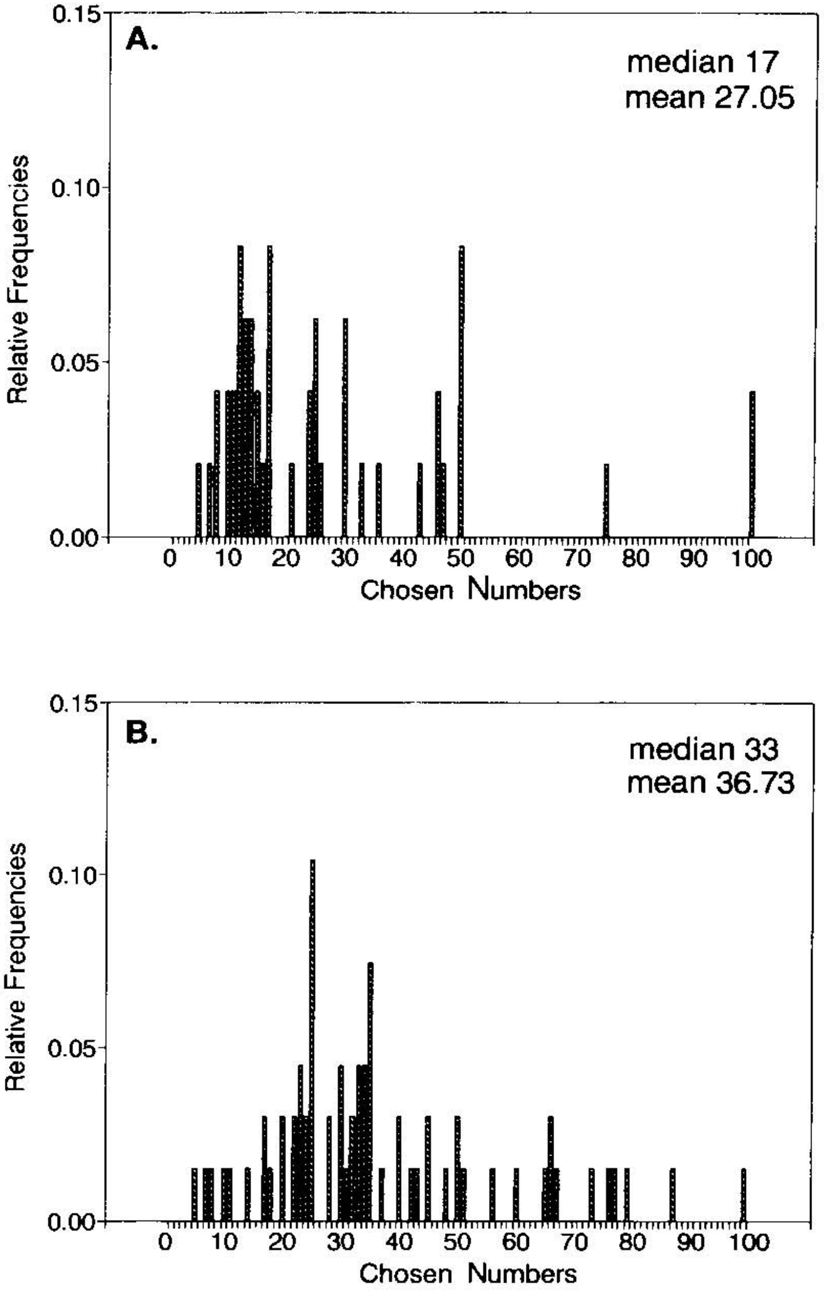
Nagel's games have a unique equilibrium, in which all guess 0; it can be found by repeatedly eliminating stupid (*dominated*, to game theorists) guesses; if *p* = 1/2, then:

● It's stupid to guess more than 50 (1/2 x 100 ≤ 50)
● Unless you think other people are stupid, it's stupid to guess more than 25 (1/2 x 50 ≤ 25)
● Unless you think other people think other people are stupid, it's stupid to guess more than 12.5 (1/2 x 25 ≤ 12.5)
● And so on, down to 6.25, 3.125, and eventually 0

The rationality-based argument for this "all-0" equilibrium is stronger than the arguments for equilibrium in the other examples, because it depends "only" on iterated knowledge of rationality, not knowledge of expectations

But even people who are rational themselves are seldom certain that others are rational, or that others believe that they themselves are rational, and so on; so they probably won't (and shouldn't) guess 0; but what do they do?

Nagel's subjects never guessed 0; their initial responses were heterogeneous, respecting 0 to 3 rounds of repeated dominance (first picture *p* = 1/2; second picture *p* = 2/3):



Boundedly rational level-*k* decision rules or "types"

Even though Nagel's subjects' initial responses deviated from equilibrium, they have a coherent structure—non-random but individually heterogeneous:

● There are spikes at $50p^k$ for k = 1,2,3—like spectrograph peaks that suggest discrete chemical elements

Similarly structured behavior patterns have been found by Stahl and Wilson '94, '95; Ho, Camerer, and Weigelt '98; Costa-Gomes, Crawford, and Broseta '01; Camerer, Ho, and Chong '04; and Costa-Gomes and Crawford '04

The data from these experiments have been analyzed by assuming that subjects' decision rules are drawn from a stable distribution of boundedly rational level-*k* or "*Lk*" types

Lk anchors its beliefs with a "naïve" prior, *L0,* and adjusts them via thought-experiments with iterated best responses:

> ● *L0* (in most applications) is random (uniformly distributed) over the set of possible decisions
> ● *L1* best responds to *L0*; thus it has a perfect model of the game but a naïve model of others
> ● *L2* (or *L3*) best responds to *L1* (or *L2*); thus they have perfect models of the game and less naïve models of others

Lk, $k > 0$, is rational in that it best responds to expectations about others, but its beliefs are based on simplified models of others that don't "close the loop" as equilibrium does:

L0 must often be adapted to the setting; but defining *Lk*, *k* > 0, by iterating best responses "works" in most settings

Lk yields a workable model of others' choices while avoiding the cognitive complexity of equilibrium; Selten ('98 *EER*):

> Basic concepts in game theory are often circular in the sense that they are based on definitions by implicit properties…. Boundedly…rational strategic reasoning seems to avoid circular concepts. It directly results in a procedure by which a problem solution is found. Each step of the procedure is simple, even if many case distinctions by simple criteria may have to be made.

In some games, *Lk* decision rules yield the same actions as equilibrium; so equilibrium predictions can be based on weaker, more plausible assumptions, and are more reliable

But in other games, *Lk* decision rules deviate systematically from equilibrium, in predictable ways

As a result, a model in which people follow a distribution of *Lk* decision rules can predict people's initial responses better than equilibrium, and yield better recommendations

A level-*k* model usually predicts a distribution of outcomes

But this uncertainty is due to the analyst's inability to observe players' types, not to players' uncertainty about each other; resemblance to mixed equilibrium is superficial

Lk types in the "scriptures"

Keynes (continuing his beauty contest quote above):

> . . . It is not a case of choosing those which, to the best of one's judgment, are really the prettiest, nor even those which average opinion genuinely thinks the prettiest. We have reached the third degree where we devote our intelligences to anticipating what average opinion expects the average opinion to be. And there are some, I believe, who practice the fourth, fifth and higher degrees.

Keynes' wording suggests finite iteration of best responses, initially anchored by players' true aesthetic preferences

(A different, social context-dependent specification of *L0*)

Benjamin Graham (of Graham and Dodd's *Security Analysis*), in *The Intelligent Investor* (thanks to Steven Scroggin for the reference):

> …imagine you are partners in a private business with a man named Mr. Market. Each day, he comes to your office or home and offers to buy your interest in the company or sell you his [the choice is yours]. The catch is, Mr. Market is an emotional wreck. At times, he suffers from excessive highs and at others, suicidal lows. When he is on one of his manic highs, his offering price for the business is high as well…. His outlook for the company is wonderful, so he is only willing to sell you his stake in the company at a premium. At other times, his mood goes south and all he sees is a dismal future for the company. In fact… he is willing to sell you his part of the company for far less than it is worth. All the while, the underlying value of the company may not have changed - just Mr. Market's mood.

Here, Graham is suggesting a best response to Mr. Market, which is a simplified model of other investors (although in context, his main goal in this passage is to keep you from becoming too emotionally involved with your own portfolio)

Thus Mr. Market is Graham's *L0* (random, though probably not uniform); so he is advocating *L1*…

But he published this, so he may actually be *L2*…

And if you ever find yourself in a situation where you need to outguess him, maybe you should be *L3*

Fiction as data? ***The Far Pavilions*** and Huarongdao

In M. M. Kaye's novel *The Far Pavilions*, the main male character, Ash, is trying to escape from his Pursuers along a North-South road; both have a single, *strategically simultaneous* choice between North and South—that is, their choices are time-sequenced, but the Pursuers must make their choice irrevocably before they learn Ash's choice

- If the pursuers catch Ash, they gain 2 and he loses 2

- But South is warm, and North is the Himalayas with winter coming, so both Ash and the Pursuers gain an extra 1 for choosing South, whether or not Ash is caught

		Pursuers	
		South (*q*)	**North**
Ash	**South (*p*)**	3 -1	0 1
	North	1 0	2 -2

Escape

(Looks almost as if Kaye borrowed from *Three Kingdoms*: Escape is just like Huarongdao…and very close to D-Day!)

Record your intuitions about what to do, as Ash or Pursuers

Escape has a unique equilibrium, in which $3p + 1(1-p) = 0p + 2(1-p)$ or $p = 1/4$, and $-1q + 1(1-q) = 0q - 2(1-q)$ or $q = 3/4$; this equilibrium is intuitive for the Pursuers, but not for Ash

But Ash chooses North and the Pursuers choose South, so the novel can continue…romantically…for 900 more pages

In equilibrium Ash North, Pursuers South has probability (1-*p*)*q* = 9/16, not bad; but try a level-*k* model with uniform *L0*

Type	Ash	Pursuers
L0	uniform random	uniform random
L1	South	South
L2	North	South
L3	North	North
L4	South	North
L5	South	South

***Lk* types' decisions in Escape**

(*Lk* types do exactly the same things in D-Day, where the Allies are analogous to Ash, and Calais to South)

Thus the level-*k* model correctly predicts the outcome provided that Ash is *L2* or *L3* and the Pursuers are *L1* or *L2*

How do we know which type Ash is? Here fiction provides data on cognition as well: Kaye recounts Ash's mentor's advice (p. 97: "ride hard for the north, since they will be sure you will go southward where the climate is kinder…)

If we take the mentor's "where" to mean "because", Ash is *L3*: Ash thinks the Pursuers are *L2*, so that the Pursuers think Ash is *L1*, so that the Pursuers think Ash thinks the Pursuers are *L0*; thus Ash thinks the Pursuers expect him to go South (because it's "kinder" and the Pursuers are no more likely to pursue him there), so Ash goes North

L3 is my record-high *k* for an *Lk* type in fiction (Poe's *The Purloined Letter* has another *L3*, but Conan Doyle doesn't even have an *L1*!); I suspect that even postmodern fiction may have no higher *Lk*s, because they wouldn't be credible

Resolving outguessing puzzles with level-*k* models

We have already seen that a level-*k* model gives a credible account of outcomes in D-Day or Huarongdao without messages, which parallel the above analysis of Escape

I conclude with level-*k* analyses of two more examples: RTH's Hide and Seek game and D-Day/Huarongdao preceded by a costless message about intentions

There are several steps in constructing a level-*k* model:

- Start with a reasonable, nonstrategic specification of *L0* (for which there is now evidence in various settings)

- Derive *Lk*'s choices for *k* = 1, 2,…, as in the above table for Escape

- Combine *Lk*'s choices with reasonable assumptions about the distribution of types in the population (for which there is a lot of evidence, which suggests that a typical population has 20-50% *L1*s and progressively smaller fractions of *L2*, *L3*, and other types)

- Sometimes, as in my analysis of D-Day/Huarongdao with a costless message about intentions, it helps to add a *Sophisticated* type, which knows everything about the game, including the distribution of *Lk* types, and plays equilibrium in a "reduced game" between *Sophisticated* players, taking the *Lk* players' choices as given

A level-*k* model of RTH's Hide-and-Seek Games
(Crawford and Iriberri '04; see paper and lecture slides at http://weber.ucsd.edu/~vcrawfor/#Hide)
Recall that in Rubinstein, Tversky, and Heller's Hide-and-Seek experiments there were strong framing effects, with *central A* most prevalent for both Hiders (37%) and Seekers (46%), and even more prevalent for Seekers than Hiders

As a result Seekers find a Treasure more than 25% of the time and have higher payoffs than in equilibrium

Puzzles:

● If Seekers are as smart as Hiders, on average, then why don't Hiders tempted to hide in *central A* realize that Seekers will be just as tempted to look there?

● Why do Hiders choose actions that, on average, allow Seekers to find them more than 25% of the time, when they could hold it down to 25% via the equilibrium mixed action (or even lower by hiding anywhere but *central A*)?

Assume that with given probabilities, each player role is filled by one of five level-*k* decision rules or "types": *L0*, *L1*, *L2*, *L3*, or *L4* (no *Sophisticated* type)

Lk, *k* > 0, anchors its beliefs in an *L0* type and adjusts via thought-experiments involving iterated best responses

Lk ignores the framing except as it influences *L0*; *L0* should reflect the simplest hypothesis a player can make about his opponent's behavior in Hide and Seek: that he will choose a salient location nonstrategically, simply because it is salient

We assume that *L0* plays A, B, A, A with probabilities *p*/2, *q*, 1–*p* –*q*, *p/*2, where *p* > 1/2 and *q* > 1/4, so *L0* favors focally labeled and/or end locations, to an equal extent for Hiders and Seekers (a uniform *L0* would replicate equilibrium)

Given this specification of *L0*:

- *L1* Hiders choose *central A* to avoid *L0* Seekers and *L1* Seekers avoid *central A* in searches for *L0* Hiders

- *L2* Hiders choose *central A* with probability between 0 and 1 and *L2* Seekers choose it with probability 1

- *L3* Hiders avoid *central A* and *L3* Seekers choose it with probability between 0 and 1

- *L4* Hiders and Seekers both avoid *central A*

With a plausible distribution of types, estimated from RTH's data (0% *L0*, 19% *L1*, 32% *L2*, 24% *L3*, 25% *L4*), the level-*k* model explains RTH's results, including the prevalence of *central A* for Hiders and Seekers and its greater prevalence for Seekers

The asymmetry in Hiders' and Seekers' behavior follows naturally from their role-asymmetric responses to *L0*, with no asymmetry in behavioral assumptions

(By contrast, "Hiders feel safer avoiding focal locations, so they are most likely to choose central A; and Seekers know this, so they are also most likely to choose central A" assumes Hiders are more sophisticated than Seekers, and doesn't explain why *central A* is more prevalent for Seekers)

A level-*k* model of D-Day/Huarongdao with a message about intentions

(Crawford *AER* '03; or see the discussion paper at http://weber.ucsd.edu/~vcrawfor/#DownLoadableDPs)

Recall that in D-Day/Haurongdao) with a costless message about intentions from the Allies/Kongming) to the Germans/ Cao Cao, all equilibria have the "sender" sending an uninformative message, which the "receiver" ignores

Given this, equilibrium with a message reduces to the mixed equilibrium of the game without a message

But the Allies took pains to fake the preparations for an invasion at Calais, and Kongming had fires lit at Huarong

In each case the sender anticipated which message would fool the receiver and chose it nonrandomly, the deception succeeded, but the sender won in the less beneficial way

- Why did the receiver allow himself to be fooled by a costless (hence easily faked) message from an *enemy*?
- Was it a coincidence that, in both Huarongdao and D-Day, the sender sent a message that fooled the receiver in a way that allowed him to win in the *less* beneficial of the two possible ways to win?
- And if the sender expected his message to fool the receiver, why didn't he reverse the message and fool the receiver in a way that allowed him to win in the *more* beneficial way? (Why didn't the Allies feint at Normandy and attack Calais? Why didn't Kongming light fires on the Main Road and ambush Cao Cao there?)

		Germans	
		Defend Calais	**Defend Normandy**
Allies	**Attack Calais**	1 -1	-2 2
	Attack Normandy	-1 1	1 -1

D-Day

Suppose the Allies' message is either "c" or "n", meaning literally (but not necessarily truthfully) that the intended invasion location is respectively Calais or Normandy

In this game with sequenced decisions, the notion of action must be extended to a contingent plan called a *strategy*

- The Allies' pure strategies are (message, action|sent message c, action|sent message n) = (c,C,C), (c,C,N), (c,N,C), (c,N,N), (n,C,C), (n,C,N), (n,N,C), or (n,N,N)

- The Germans' pure strategies are (action|received message c, action|received message n) = (N,N), (N,C), (C,N), or (C,C)

Allies' and Germans' types are drawn from separate distributions, including both boundedly rational, or *Mortal*, types and a strategically rational, or *Sophisticated*, type

Sophisticated types know everything about the game, including the distribution of *Mortal* types; and play equilibrium in a "reduced game" between *Sophisticated* players, taking *Mortal* players' choices as given

Mortal types' behaviors regarding the message are anchored on analogs of *L0* based on truthfulness or credulity, as in the informal literature on deception:

- *W0* ("wily") for senders (*Mortal* Allies)

- *S0* ("skeptical") for receivers (*Mortal* Germans)

Higher-level *Mortal* types are defined by iterating best responses to *W0* or *S0*, just as *Lk* best responds to *Lk-1*:

- Higher-level *Mortal* Allies are *Liars* or *Truthtellers*

- Higher-level *Mortal* Germans are *Believers* or *Inverters*

Thus in the history, if the Allies were *Mortal* rather than *Sophisticated*, then the Allies were *Liars*, who expected the Germans to be deceived by their false message—not because the Germans were credulous, but because they were *Believers*, who would invert it one too many times

But if Kongming was *Mortal*, then he was a *Truthteller*, who expected Cao Cao, as an *Inverter*, to be deceived by a truthful message (but this difference is inessential here)

Mortal Allied types, *Wk* for *k* > 1, always expect to fool the Germans, either by lying (like the Allies did) or by telling the truth (like Kongming did)

Given this, all *Mortal* Allied types *Wk* for *k* > 1 send a message that they expect to make the Germans think they will attack Normandy; and then attack Calais instead

If we knew the Allies and Germans were *Mortal*, we could now derive the model's implications from an estimate of the type frequencies; but the analysis can usefully be extended to allow the possibility of *Sophisticated* Allies and Germans

To do this, we plug in the distributions of *Mortal* Allies' and Germans' behavior to obtain the reduced game between *Sophisticated* Allies and Germans, and study its equilibria

There are two cases, with different implications:

- When *Sophisticated* Allies and Germans are common—not the most plausible case—then the reduced game has a mixed equilibrium whose outcome is virtually equivalent to D-Day's without communication

- When *Sophisticated* Allies and Germans are rare, the game has an essentially unique pure equilibrium, in which *Sophisticated* Allies send the message that fools the most common *Mortal* Germans, *Believer* or *Inverter*, and then attack Normandy; and *Sophisticated* Germans defend Calais

In this equilibrium the Allies' message and action are part of a single, integrated strategy; and the probability of attacking Normandy is much higher than if no message was possible

Note that there is no pure equilibrium in which *Sophisticated* Allies feint at Normandy and attack Calais (though this has positive probability in mixed equilibria)

Thus for plausible parameter values the model "explains" the history with equally *Sophisticated* Allies and Germans

Conclusion

In this lecture I have compared history, data, or intuitions about strategic behavior with equilibrium predictions in some simple examples of outguessing games, highlighting puzzles that equilibrium either does not address, or gets wrong

I then described a non-equilibrium model of initial responses to games based on "level-*k*" thinking, which is closer to strategic intuition and experimental evidence, and compared it with equilibrium as a model of behavior in novel games

In some games the level-*k* model predicts the same actions as equilibrium, so equilibrium predictions can be based on weaker, more plausible assumptions, and are more reliable

In other games the level-*k* model's predictions deviate systematically from equilibrium, but in predictable ways

In outguessing games the level-*k* model deviates systematically from equilibrium, and these deviations bring its predictions closer to evidence and intuition, resolving some of the puzzles left open by equilibrium analysis