Outguessing and Deception in Novel Strategic Situations

Vincent P. Crawford, University of California, San Diego
MEDS, Northwestern University, 4 October 2005

Many strategic situations in business, politics, security, or war are well approximated by *outguessing games*—pure-conflict situations in which some players want to match others' actions (in some commonly understood sense) and others want to avoid matching:

- Entry games where entry requires a differentiated product and blocking it requires matching the entrant's design (e.g. Roger Myerson's Ware Case; see Bob Weber's course materials at http://www.kellogg.northwestern.edu/faculty/weber/decs-452/)

- Election campaigns in which a challenger can win only by campaigning in different areas than the incumbent

- Hide-and-seek games like those that underlie the quotes:

  "Any government wanting to kill an opponent…would not try it at a meeting with government officials."
  —on the poisoning of Ukrainian presidential candidate Viktor Yushchenko in 2004

  "…in Lake Wobegon, the correct answer is usually 'c'."
  —Garrison Keillor on multiple-choice tests

- Games of military strategy like D-Day (Crawford (AER 2003)):

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

D-Day (as Perturbed Matching Pennies)

Here 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

Deception

In outguessing games strategic deception via approximately costless signaling of intentions also plays an important role (e.g. the Ware Case, the Allies' Operation Fortitude South preceding D-Day)

Thus I will also consider outguessing games preceded by a one-way "cheap-talk" message about intentions

Equilibrium in outguessing games

Nash equilibrium makes very clear predictions about behavior in outguessing games and about strategic deception

These are often accurate when players have enough experience with analogous games to learn to predict others' responses (e.g. Walker and Wooders, "Minimax Play at Wimbledon" (AER 2001))

But in novel strategic situations there may be no analogous games

Equilibrium must then come from thinking rather than learning, and its predictions are correspondingly less reliable for initial responses

(If investigators believed that "any government wanting to kill an opponent…would not try it at a meeting with government officials," then that is precisely where a government *would* want to try it…yet this non-equilibrium principle is ubiquitous in "folk game theory")

Outline

The talk begins by using examples to highlight the strategic issues any successful theory of outguessing and deception must address

I then compare equilibrium predictions in the examples with history, experimental data, or intuitions about strategic behavior, highlighting behavioral puzzles left open by equilibrium analysis

I then introduce a structural non-equilibrium model of initial responses based on "level-*k* thinking" that is based on recent experimental work

In some games a level-*k* model's predictions coincide with equilibrium, in which case equilibrium predictions rest on weaker behavioral assumptions and are correspondingly more reliable

In other games a level-*k* model's predictions deviate systematically from equilibrium, and using a level-*k* model to predict the deviations can help to resolve empirical puzzles

Today I consider games with payoff asymmetries like D-Day, with and without preplay communication about intentions

Tomorrow I will consider hide-and-seek games with non-neutral framing of locations, as in the Yushchenko quote

Both talks are based partly on joint work with Miguel Costa-Gomes of York, Bruno Broseta of Red de Institutos Tecnológicos de la Comunidad Valenciana, and Nagore Iriberri of UCSD

Equilibrium analysis of 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 unperturbed Matching Pennies (no 2's, all 1's), where the equilibrium p and q = 1/2

Record your intuitions about how to play as Allies, as Germans

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

D-Day

In a Nash equilibrium, each player chooses his best action, given correct (probabilistic) expectations about the other's action

If players choose deterministically between Calais and Normandy, then D-Day has no equilibrium

But in D-Day it is important to be unpredictable, so it is natural to allow randomized (*mixed*) as well as unrandomized (*pure*) actions

Mixed actions are less weird than they may seem because a player's action need only be unpredictable to others, not himself: we can view the equilibrium as an equilibrium in *beliefs*

The equilibrium mixed strategies 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 may match your qualitative intuition for Germans because their better-on-average action, Defend Calais, has $q > 1/2$

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

D-Day's equilibrium *must* be counterintuitive because if Allies tried to exploit the ease of attacking Calais in an obvious way ($p = 1$), and this was predictable, then Germans could neutralize them 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 $p < 1/2$; $p = 2/5$ yields them payoff 1/5, more than equilibrium in Matching Pennies

This principle seems too subtle to be identified in bridge textbooks or informal writing on strategy (but see vN-M (1953); vN (1953); Crawford and Smallwood, Theory and Decision (1984))

Nonetheless, people (experimental subjects) systematically respond to the asymmetries, in ways that deviate from equilibrium

E.g. perturbed Matching Pennies example from Camerer talk slides at http://www.hss.caltech.edu/~camerer/SS200/bgtheory05.ppt (see also Rosenthal, Shachat and Walker (IJGT 2003)):

	Row step thinker choices							CH (τ = 1.62)	mixed	
	L	R	0	1	2	3	4…	pred'n	equilm	data
T	2,0	0,1	.5	1	1	0	0	.68	.50	.72
B	0,1	1,0	.5	0	0	1	1	.32	.50	.28
0	.5	.5								
1	.5	.5								
2	0	1								
3	0	1								
4	0	1								
5	0	1								
CH	.26	.74								
mixed	.33	.67								
data	.33	.67								

[τ (roughly the average *k* below) = 1.62 makes their Cognitive Hierarchy model close to the level-*k* model proposed below; the predictions would be somewhat different, but still closer to the data than equilibrium]

D-Day puzzle

How should we advise people to respond to payoff asymmetries like those in D-Day?

Equilibrium makes a precise prediction about such responses, but (under the plausible equilibrium in beliefs interpretation) says it doesn't matter what you do, beyond avoiding dominated actions

If these people aren't playing equilibrium, what are they doing?

And how should we advise people to play against them?

Huarongdao

It is interesting to consider an ancient Chinese antecedent of D-Day's, Huarongdao, in which General Cao Cao chooses between two roads, trying to avoid capture by General Kongming

(Huarongdao adds a second data point to the D-Day observation in my AER 2003 paper; thanks to Duozhe Li of CUHK 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 a unique mixed equilibrium

D-Day and Huarongdao with costless message about intentions

Now give the Allies, and Kongming, an opportunity to send a message about intentions before the actions are chosen…

as in Kongming's fires along the Huarong road and the Allies' faked preparations for invasion at Calais in Operation Fortitude

An Inflatable "Tank" from Operation Fortitude

In such games, with approximately costless messages, all equilibria have the sender sending an uninformative message, which the receiver ignores

Otherwise the receiver would benefit by responding to the message; but in a 0-sum game such a response would hurt the sender, who would do better to make his message uninformative

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

But deceptive signals about one's intentions are ubiquitous in outguessing games; and in both D-Day and Huarongdao:

- The sender anticipated which message would fool the receiver and chose it nonrandomly

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

- The deception succeeded, but the sender won in the less beneficial of the two possible ways

(An unimportant difference: in D-Day the message was literally deceptive but the Germans "believed" it, either because they were credulous or because they inverted it one too many times; while Kongming's message was truthful—he lit fires on the Huarong Road and ambushed Cao Cao there—but Cao Cao inverted it

One advantage of fiction as data is that it can reveal cognition:

- *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'?"

- It also gives Cao Cao's rationale for inverting the message: "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….)

D-Day/Huarongdao with messages puzzle

How should we advise people to send messages or read others' messages, and to play the underlying game, in games like D-Day/Huarongdao with a costless message about intentions?

Equilibrium again makes a precise but unhelpful prediction

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 two possible ways?

(If he expected his message to fool the receiver, why didn't he reverse it and fool the receiver in a way that allowed him to win in the *more* beneficial way? I.e. why didn't the Allies feint at Normandy and attack at Calais? Why didn't Kongming light fires on the Main Road and ambush Cao Cao in the more convenient location?)

If these people aren't playing equilibrium, what are they doing?

And how should we advise people to play against them?

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

Rationalizability is even less helpful than equilibrium here; we need some way to model non-equilibrium strategic 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 some of the puzzles left open by equilibrium

Consider subjects' initial responses in Nagel's AER 1995 "guessing" or "beauty contest" (Keynes quote below) games:

● 15-18 subjects simultaneously guess between [0,100]

● The subject whose guess is closest to a target 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 (or more politely, *dominated*) guesses

If p = 1/2, then:

● It's stupid to guess more than 50 (1/2 x 100 ≤ 50)

● Unless you think the other people are stupid, it's stupid to guess more than 25 (1/2 x 50 ≤ 25)

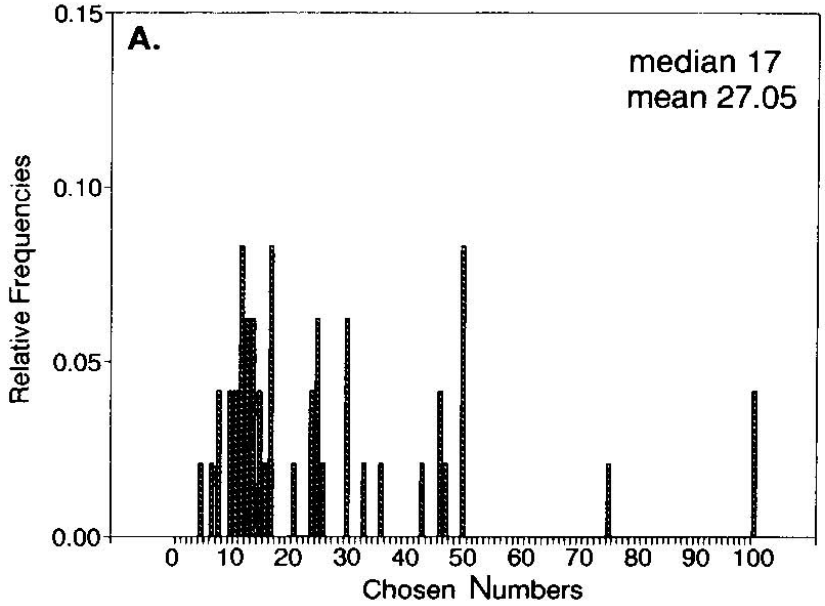
● Unless you think the other people think the other people are stupid, it's stupid to guess more than 12.5 (1/2 x 25 ≤ 12.5)

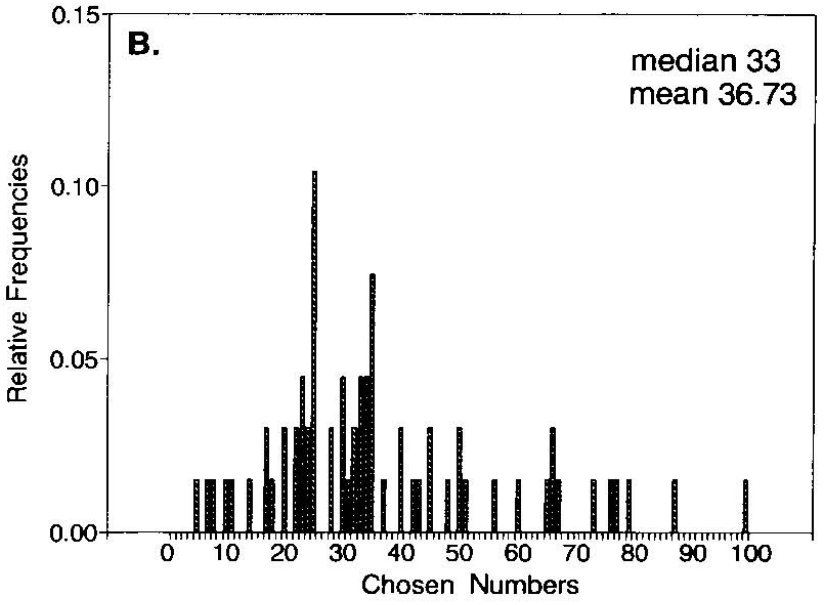
● And so on, down to 6.25, 3.125, and eventually 0

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 infinitely 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 iterated dominance (Figure 1 in her paper; first picture p = 1/2; second picture p = 2/3):





First picture p = 1/2; second picture p = 2/3

Level-*k* decision rules or "types"

Even though Nagel's subjects' initial responses deviated from equilibrium, their responses have a coherent, non-random, and individually heterogeneous structure: there are spikes at $50p^k$ for k = 1,2,3—like the spectrograph peaks of discrete chemical elements

Similar patterns have been found by Stahl and Wilson (JEBO 1994, GEB 1995); Ho, Camerer, and Weigelt (AER 1998); Costa-Gomes, Crawford, and Broseta (EMT 2001); Camerer, Ho, and Chong (QJE 2004); Costa-Gomes and Crawford (2004); and Crawford and Iriberri (2005a, 2005b)

The data from these experiments have been analyzed mainly 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, nonstrategic prior *L0,* and adjusts them via thought-experiments with iterated best responses:

- *L0* is most often taken to be uniform random 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

L0 must often be adapted to the setting (as in the games with communication below and hide-and-seek games tomorrow); but defining *Lk*, $k > 0$, by iterating best responses works in most settings

Lk, $k > 0$, is rational in that it understands the structure of the game and best responds to beliefs about others' decisions

It differs from equilibrium in that its beliefs are based on simplified models of others that don't "close the loop" as equilibrium does

This yields a workable model of others' choices while avoiding the cognitive complexity of equilibrium; Selten (EER 1998):

> 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.

The estimated type frequencies are reasonably stable across different settings, with significant weight only on *L1*, *L2*, and *L3*

In some games the empirically significant *Lk* types' predictions coincide with equilibrium, in which case equilibrium predictions rest on weaker behavioral assumptions and are correspondingly more reliable

In other games *Lk* types' predictions deviate systematically from equilibrium

In such games a model in which people follow a distribution of *Lk* rules can predict initial responses better than equilibrium

Although the model usually predicts a distribution of outcomes, the resemblance to mixed equilibrium is superficial (although the randomness is real to players' opponents)

Lk types in the "scriptures"

Imagine you are an investor deciding what to do when the NYSE reopens after 9/11: Do you dump airline stocks in response to the new information about airlines' likely future profitability, or do you wait and try to profit from others' overreaction to this information?

(This setting is more complex than those considered so far because of incomplete information; Crawford and Iriberri (2005b) take a first step in this direction by analyzing "Level-*k* Auctions")

Keynes' famous "beauty contest" example (*The General Theory*, ch. 12), which inspired Nagel's experiment, likens 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. 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 here 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 of UCSD):

> …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* (but not higher: it can be just as bad to be too sophisticated as to be too unsophisticated)

Costa-Gomes and Crawford's (2004) guessing experiments

People often assume Nagel's spikes are evidence of finitely iterated dominance, and this is not separated from level-*k* types in her design

But Costa-Gomes and Crawford (2004) (see also Costa-Gomes, Crawford, and Broseta (EMT 2001)) (http://weber.ucsd.edu/~vcrawfor/#Guess) separate them and show that the data overwhelmingly favor the level-*k* interpretation

Costa-Gomes and Crawford (2004) also show that one can explain a large fraction of the deviations from equilibrium using a level-*k* model, and that nothing else explains a significant fraction

Their results are specific to a particular context, but their design has considerable power to discriminate among alternative interpretations of subjects' behavior; and to the extent that one can check, the results are consistent with previously obtained results from other settings

The level-*k* model the results suggest provides a simple, tractable alternative to equilibrium models of initial responses

The clarity of the data should help to allay the fear that if we don't assume equilibrium, "anything can happen"

Costa-Gomes and Crawford's (2004) Design

Game-theoretically naïve subjects played 16 different two-person guessing games; anonymously, randomly paired with no feedback during play to suppress learning and repeated-game effects

● In each game, two players make simultaneous guesses

● Each player has a lower and an upper limit, both positive, so games are finitely dominance-solvable

● But players are not required to guess between their limits: instead guesses outside the limits are *automatically* adjusted up to the lower limit or down to the upper limit as needed (a trick to enhance separation of types' search implications)

● Each player has a target, and his payoff increases the closer his adjusted guess is to his target times the other's adjusted guess

● The targets and limits vary independently across players and games, with targets both < 1, both > 1, or mixed (equilibrium is determined by players' lower (upper) limits when the product of their targets is < (>) 1, which allows additional inferences about cognition)

● Because the targets and limits vary, subjects don't know them

● Costa-Gomes and Crawford presented the games with targets and limits hidden, giving subjects free access to them *game by game*, publicly announcing all other aspects of the structure; this allowed a study of subjects' cognition by monitoring their information searches

● Low search costs then made the games' structures effectively public knowledge, so that (with suppression of learning and repeated-game effects) the design induced a series of 16 independent complete-information games

Design yields strong separation of leading types, very clear results:

Types' guesses in the 16 games, in (randomized) order played

	L1	*L2*	*L3*	*D1*	*D2*	*Eq.*	*Sop.*
1	600	525	630	600	611.25	750	630
2	520	650	650	617.5	650	650	650
3	780	900	900	838.5	900	900	900
4	350	546	318.5	451.5	423.15	300	420
5	450	315	472.5	337.5	341.25	500	375
6	350	105	122.5	122.5	122.5	100	122
7	210	315	220.5	227.5	227.5	350	262
8	350	420	367.5	420	420	500	420
9	500	500	500	500	500	500	500
10	350	300	300	300	300	300	300
11	500	225	375	262.5	262.5	150	300
12	780	900	900	838.5	900	900	900
13	780	455	709.8	604.5	604.5	390	695
14	200	175	150	200	150	150	162
15	150	175	100	150	100	100	132
16	150	250	112.5	162.5	131.25	100	187

Subjects with Types Identifiable from Guesses

Type	*L1*	*L2*	*L3*	*Eq.*
# of type's	16,16	16,16	9,11	11,12
w/in 0, w/in 25	15,15	16,16	7,11	11,11
	15,15	15,15	7,10	9,10
	14,14	13,13		8,10
	14,14	13,13		8,8
	14,14	11,14		8,8
	13,13	11,12		7,8
	12,12	11,11		7,7
	10,11	9,10		6,7
	10,10	8,10		5,5
	10,10	8,8		
	10,10	7,8		
	10,10	6,11		
	9,9	6,8		
	8,9	6,8		
	7,8			
	3,12			

On average 90% of subjects' guesses respected simple dominance, much more than random (~60%) and typical of initial responses

All but 12 respected dominance in 13 or more games (80%), suggesting that they understood the games and maximized self-interested expected payoffs, given coherent beliefs

43 of 88 subjects made 7-16 of some type's exact (within 0.5) guesses: far more than could occur by chance, given the strong separation of types' guesses and the fact that guesses could take from 200 to 800 different rounded values

But 35 of those 43 subjects conformed closely to types other than *Equilibrium*: 20 to *L1*, 12 to *L2*, and 3 to *L3*

Given our type definitions, those subjects' deviations from equilibrium can be confidently ascribed to non-equilibrium beliefs rather than altruism, spite, confusion, or irrationality

(The results for guesses also favor CGCB's noiseless definition of *Lk*, *k* > 1, over SW's, which best responds to a noisy *Lk-1*; and provide evidence against types that depend on estimated population parameters, such as SW's *Worldly*)

Studying cognition via guesses and information search

Costa-Gomes and Crawford linked guesses and information search by assuming that each subject has a single, "pure" type, which determines both guesses and search in the 16 games

The types *L1*, *L2*, *L3*, *D1*, *D2*, *Equilibrium*, and *Sophisticated* provide a kind of basis for the enormous space of possible guesses and searches, imposing enough structure to make it meaningful to ask if they are related in a coherent way

They derived types' search implications as follows:

● Standard assumptions imply that a type will look up all freely available information that might affect its guess

● Each type is naturally associated with algorithms that describe how to process this information into a guess

● They used a type's algorithms as models of cognition, and derive the search implications of those algorithms under conservative assumptions about how cognition affects search (Table XI; assumptions are needed because if a subject memorized parameters, look-up order could be unrelated to cognition)

Subjects' searches generally reaffirm their type estimates based on guesses alone

In the end 52 of 88 subjects are reliably identified: 27 as *L1*, 13 as *L2*, 10 as *Equilibrium*, and possibly one each as *L3* or *Sophisticated*

Table XI. Selected Baseline Subjects' Information Searches and Estimated Types' Search Implications

MouseLab box numbers

	A	*b*	*p*
You (*i*)	1	2	3
S/he (*j*)	4	5	6

Types' Search Implications

L1	{[4,6],2}
L2	{([1,3],5),4,6,2}
L3	{([4,6],2),1,3,5}
D1	{(4,[5,1], (6,[5,3]),2}
D2	{(1,[2,4]),(3,[2,6]),(4,[5,1],(6,[5,3]),5,2}
Eq	{[2,5],4} if pr. tar.<1,{[2,5],6} if > 1

Subject	**101**	**118**	**413**	**108**	**206**	**309**	**405**	**210**	**302**	**318**	**417**	**404**	**202**	**310**	**315**
Type(#rt.)	L1 (15)	L1 (15)	L1 (14)	L2 (13)	L2 (15)	L2 (16)	L2 (16)	L3 (9)	L3 (7)	L1 (7)	Eq (8)	Eq (9)	Eq (8)	Eq (11)	Eq (11)
Alt.(#rt.)								Eq (9)	Eq (7)	D1 (5)	L3 (7)	L2 (6)	D2 (7)		
Alt.(#rt.)								D2 (8)			L2 (5)		L3 (7)		
Est. style	early/late	early	late	early	early	early/late	early	early	early	early	early	early	early	early/late	early
Game															
1	146246213	24613462624132*135	1234565456123463*	135642	53314621 3	1352	14465231331254623212512	123456123456213456254213654	221135465645213213 45456*541	132456465252 13242*1462	252531464656446531641252462121 3	462135464655645515 21354*135462426256356234131354645	123456254613621342*525	123126544121565421254362*21545 4*	213465624163564121325466
2	46213	2462622131	12356462213*	1356423	53146231	1352631526*2*3	132456253156456545463123156562 62	123456465562231654456*2	213546566213545463 21*26654123	132465132*462	25523662*365243563	462461352524261315463562	123456445613255462513565 23	123546216326231456*62	134652124653656121 3
3	462*46	246242466413*426	264231	13564253	53516422231	135263	3124565231*1236545 5233**513	1234556456123	265413232145 563214563214 523*654123	1346521323*4	52136364152652 63*652	462135215634*52	1234561235623	123655463213	132465544163*3625

More fiction as data: Level-*k* thinking in *The Far Pavilions*, Huarongdao, and D-Day

Early in M. M. Kaye's novel *The Far Pavilions*, the main male character, Ash, tries to escape from his Pursuers along a North-South road; both have a single, strategically simultaneous choice between North and South—their choices are time-sequenced, but the Pursuers must choose 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

Escape

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

(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 random 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 if Ash is *L2* or *L3*? Fiction reveals cognition through his mentor's advice: "ride hard for the north, since they will be sure you will go southward where the climate is kinder…" (p. 97)

If we read "where" as "because," Ash is *L3*: Ash thinks the Pursuers are *L2*, and so thinks the Pursuers think Ash is *L1*, and so thinks 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

The Pursuers are probably *L2* (but they have no mentor to tell us)

L3 is my record *k* for an *Lk* type in fiction (Poe's story *The Purloined Letter* also has an *L3* (http://weber.ucsd.edu/~vcrawfor/#Hide); Conan Doyle doesn't even have an *L1*...even postmodern fiction may have no higher *Lk*s, perhaps because they wouldn't be credible

If we were doing the analysis without an omniscient narrator, we could estimate that a typical population of Pursuers (not too bright) may have 30-50% *L1*s and progressively fewer *L2*, *L3*, etc.

Thus (consulting the table) the Pursuers are quite likely to go South, and Ash's choice of North is pretty robustly optimal

A similar analysis yields similar conclusions in games like D-Day or Huarongdao without messages, much as in Camerer's analysis of the perturbed Matching Pennies game displayed above

A level-*k* model of D-Day/Huarongdao with costless messages

I conclude by sketching a level-*k* analysis of D-Day/Huarongdao with a costless Allied message about intentions (Crawford, (AER 2003))

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

D-Day

Assume Allies' and Germans' types are drawn from separate distributions, including both boundedly rational, or *Mortal*, types and a strategically rational, or *Sophisticated*, type (interesting but rare)

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

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

- *W0* ("wily") for senders (*Mortal* Allies) tells the truth
- *S0* ("skeptical") for receivers (*Mortal* Germans) believes whatever it is told

Suppose the Allies' message is "c" or "n", meaning literally (but not necessarily truthfully) that the intention is Calais or Normandy

Extend the notion of action 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)

Derive Higher-level *Mortal* types *Wk*'s and *Sk*'s choices for *k* = 1, 2,…, as in the table for Escape or Table 1 in Crawford (AER 2003):

Sender type	**Behavior (b.r. ≡ best response)**	**message, action|sent u, action|sent d**
Credible ≡ *W0*	tells the truth	u,U,D
***W1* (Wily)**	lies (b.r. to *S0*)	d,D,U
W2	tells truth (b.r. to *S1*)	u,U,D
W3	lies (b.r. to *S2*)	d,D,U
Sophisticated	b.r. to population	depends on the type probabilities
Receiver type	**Behavior**	**action|received u, action|received d**
Credulous ≡ *S0*	believes (b.r. to *W0*)	R, L
***S1* (Skeptical)**	inverts (b.r. to *W1*)	L, R
S2	believes (b.r. to *W2*)	R, L
S3	inverts (b.r. to *W3*)	L, R
Sophisticated	b.r. to population	depends on the type probabilities

Table 1. Plausible *Mortal* and *Sophisticated* sender and receiver types

Mortal types, like other boundedly rational types, use step-by-step procedures that generically determine unique, pure strategies, avoid simultaneous determination of the kind used to define equilibrium

A *Wily* Sender/Ally, *Wj*, with *j* odd always lies; lump these *Mortal* sender types together under the heading *Liars*

A *Wily* sender/Ally, *Wj*, with *j* even (including *Credible* as honorary *Wily* type, *W0*) always tells the truth; lump these *Mortal* sender types together as *Truthtellers*

A *Skeptical* receiver/German, *Sk*, with *k* odd always inverts the sender's message, and with *k* even (including *Credulous* as *S0*) always believes it; lump these *Mortal* receiver types together as *Inverters and Believers*

(If the Allies were *Mortal* rather than *Sophisticated*, then they 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)

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

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 type frequencies

But the analysis can usefully be extended to allow the possibility of *Sophisticated* Allies and Germans

To do this, note first that *Mortals'* strategies are determined independently of each other's and *Sophisticated* players' strategies, and so can be treated as exogenous (but they affect others' payoffs)

Then plug in the distributions of *Mortal* Allies' and Germans' independently determined behavior to obtain a "reduced game" between possibly *Sophisticated* Allies and Germans

Because *Sophisticated* players' payoffs are influenced by *Mortal* players' decisions, the reduced game is no longer zero-sum, its messages are not cheap talk, and it has incomplete information

(The sender's message, which is ostensibly about his intentions, is in fact read by the receiver as a signal of his type)

The equilibria of the reduced game are determined by the population frequencies of *Liars*, *Truthtellers*, and *Sophisticated* senders, and of *Believers*, *Inverters*, and *Sophisticated* receivers

There are two leading cases, with different implications:

- When *Sophisticated* Allies and Germans are common—not that plausible—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 can predict *Sophisticated* Germans' action, and vice versa; and 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 (there is no pure equilibrium in which *Sophisticated* Allies feint at Normandy and attack Calais (though this outcome has positive probability in a mixed equilibrium))

In the pure 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 about intentions was possible

The Allies choose their message nonrandomly, the deception succeeds (most of the time), but it allows the Allies to win in the less beneficial of the possible ways

Thus for plausible parameter values, without postulating an unexplained difference in the sophistication of Allies and Germans, the model explains why the Germans allowed themselves to be "fooled" by a costless message from an enemy, and why the Allies didn't feint at Normandy and attack Calais

More details

A *Sophisticated* receiver's strategy is R,R in *all* pure-strategy sequential equilibria because if a sender deviates from his pure-strategy equilibrium message, it "proves" that sender is *Mortal,* making receiver's best response R; but in the only pure-strategy equilibria in which a *Sophisticated* receiver's strategy is *not* R,R, a *Sophisticated* sender plays U on the equilibrium path, so a *Sophisticated* receiver must also play R on the equilibrium path

Because a *Sophisticated* sender cannot truly fool a *Sophisticated* receiver in equilibrium, whichever action he chooses in the underlying game, it is always best to send the message that fools whichever type of *Mortal* receiver, *Believer* or *Inverter*, is more likely

The only remaining choice is whether to play U or D, when, with the optimal message, the former action fools $\max\{r_b,r_i\}$ *Mortal* receivers at a gain of *a* per unit and the latter fools them at a gain of 1 per unit, but also "fools" r_s *Sophisticated* receivers; simple algebra reduces this question to whether $a \max\{r_b,r_i\} + \min\{r_b,r_i\} > 1$ or < 1

(There are also hybrid mixed equilibria when a *Sophisticated* sender (receiver) has high (low) probability, in which randomization is confined to the sender's message, and "punishes" a *Sophisticated* receiver for deviating from R,R in a way that allows the sender to realize higher expected payoff; these equilibria are like the pure-strategy equilibria for adjoining parameter configurations, and converge to them as the relevant population parameters converge)

Conclusion

In this lecture I have considered some simple examples of outguessing games with and without preplay communication about intentions, focusing on initial rather than learned responses to the games

I then compared history, data, and intuitions about strategic behavior with equilibrium predictions in the examples, highlighting puzzles that equilibrium either does not address, or gets wrong

I then described a structural non-equilibrium model of initial responses to games based on "level-*k*" thinking, which is closer to strategic intuition and experimental evidence

In some games a level-*k* model's predictions coincide with equilibrium, in which case equilibrium predictions rest on weaker behavioral assumptions and are correspondingly more reliable

In other games, including the outguessing games considered here, a level-*k* model's predictions deviate systematically from equilibrium

In outguessing games level-*k* models' deviations bring their predictions closer to evidence and intuition, resolving some empirical puzzles