Deceived while Rational?
Game-theoretic Models of Deception and Gullibility

Vincent P. Crawford

**University of Oxford, All Souls College,
and University of California, San Diego**

(Based in part on "Modeling Strategic Communication: From Rendezvous and Reassurance to Trickery and Puffery," Nancy Schwartz Memorial Lecture, 2017)

UC San Diego

European Research Council

Established by the European Commission

Revised 31 May 2019

Deceived while Rational?
Game-theoretic Models of Deception and Gullibility

Vincent P. Crawford

**University of Oxford, All Souls College,
and University of California, San Diego**

(Based in part on "Modeling Strategic Communication: From Rendezvous and Reassurance to Trickery and Puffery," Nancy Schwartz Memorial Lecture, 2017)

"Any fool can tell the truth, but it requires some sense to know how to lie well."—*The Notebooks of Samuel Butler*, 1912

UC San Diego

European Research Council
Established by the European Commission

Revised 31 May 2019

Three puzzles game theory should resolve: Trickery, Puffery, Gullibility

Trickery: Animarumadversion (resemblance to actual Fellows is coincidental)

Professor Gamma thinks a review by Distinguished Fellow Alpha of Gamma's thousand-page critique of economics, *Capitalist Apologetics in the 21st Century*, would be just the thing to launch it into the neo-post-neoliberal stratosphere

Alpha secretly thinks the book is destined to become space junk, and that even Gamma's mother would be hard-pressed to review it favorably

Three puzzles game theory should resolve: Trickery, Puffery, Gullibility

Trickery: Animarumadversion (resemblance to actual Fellows is coincidental)

Professor Gamma thinks a review by Distinguished Fellow Alpha of Gamma's thousand-page critique of economics, *Capitalist Apologetics in the 21st Century*, would be just the thing to launch it into the neo-post-neoliberal stratosphere

Alpha secretly thinks the book is destined to become space junk, and that even Gamma's mother would be hard-pressed to review it favorably

Alpha and Gamma each plan to dine once this weekend, on Saturday or Sunday

It's known that at dinner, Alpha lacks the will to refuse even odious favors

It's also known that Alpha prefers dining on Sundays, and Gamma on Saturdays

Three puzzles game theory should resolve: Trickery, Puffery, Gullibility

Trickery: Animarumadversion (resemblance to actual Fellows is coincidental)

Professor Gamma thinks a review by Distinguished Fellow Alpha of Gamma's thousand-page critique of economics, *Capitalist Apologetics in the 21st Century*, would be just the thing to launch it into the neo-post-neoliberal stratosphere

Alpha secretly thinks the book is destined to become space junk, and that even Gamma's mother would be hard-pressed to review it favorably

Alpha and Gamma each plan to dine once this weekend, on Saturday or Sunday

It's known that at dinner, Alpha lacks the will to refuse even odious favors

It's also known that Alpha prefers dining on Sundays, and Gamma on Saturdays

It is 9:57 Saturday morning: Just time for Alpha to respond to Gamma's email asking about his dinner plans, and then email the Lodge

Alpha emails Gamma saying that he "hopes" to dine on Sunday

Should Gamma book his dinner for Saturday or Sunday?

Three puzzles game theory should resolve: Trickery, Puffery, Gullibility

Trickery: Animarumadversion (resemblance to actual Fellows is coincidental)

Professor Gamma thinks a review by Distinguished Fellow Alpha of Gamma's thousand-page critique of economics, *Capitalist Apologetics in the 21st Century*, would be just the thing to launch it into the neo-post-neoliberal stratosphere

Alpha secretly thinks the book is destined to become space junk, and that even Gamma's mother would be hard-pressed to review it favorably

Alpha and Gamma each plan to dine once this weekend, on Saturday or Sunday

It's known that at dinner, Alpha lacks the will to refuse even odious favors

It's also known that Alpha prefers dining on Sundays, and Gamma on Saturdays

It is 9:57 Saturday morning: Just time for Alpha to respond to Gamma's email asking about his dinner plans, and then email the Lodge

Alpha emails Gamma saying that he "hopes" to dine on Sunday

Should Gamma book his dinner for Saturday or Sunday?

Alpha then emails the Lodge, booking his dinner for…Saturday

Trickery: D-Day

(Framing example from Crawford, 2003 *American Economic Review*; I admit that the actual history was much more complex)

A "Tank" from Operation Fortitude South in the Thames Estuary, 1944

The Allies choose where to invade Europe on D-Day: Calais or Normandy

- Attacking an undefended Calais is better for the Allies than attacking an undefended Normandy (but the Germans know that too)

- Defending an unattacked Normandy is worse for the Germans than defending an unattacked Calais

The Allies choose where to invade Europe on D-Day: Calais or Normandy

- Attacking an undefended Calais is better for the Allies than attacking an undefended Normandy (but the Germans know that too)

- Defending an unattacked Normandy is worse for the Germans than defending an unattacked Calais

Before the attack, the Allies place a fake invasion army in the Thames Estuary: an approximately cheap talk message regarding their intentions, with an obvious literal meaning (http://en.wikipedia.org/wiki/Operation_Fortitude)

The Germans believe the message (or perhaps invert it one too many times) and overdefend Calais; the Allies invade at Normandy

Trickery: Huarongdao (from Luo Guanzhong's novel, *Three Kingdoms*; Crawford, Costa-Gomes, and Iriberri, 2013 *Journal of Economic Literature*)

Defeated, fleeing General Cao Cao chooses between two escape routes, the easy Main Road and the awful Huarong Road, trying to evade pursuing General Kongming (http://chinesepuzzles.org/huarong-pass-sliding-block-puzzle/)

Other things equal, both generals prefer the Main Road

Trickery: Huarongdao (from Luo Guanzhong's novel, *Three Kingdoms*; Crawford, Costa-Gomes, and Iriberri, 2013 *Journal of Economic Literature*)

Defeated, fleeing General Cao Cao chooses between two escape routes, the easy Main Road and the awful Huarong Road, trying to evade pursuing General Kongming (http://chinesepuzzles.org/huarong-pass-sliding-block-puzzle/)

Other things equal, both generals prefer the Main Road

Kongming waits in ambush along Huarong Road and sets campfires there, sending an approximately cheap talk message with an obvious literal meaning

Cao Cao expects a lie, inverts the message, and is caught on Huarong Road

Puzzles

In all three examples (two of them nearly real), the message-sender deceived the message-receiver and won, but in the *less* beneficial of the two possible ways

- Why did the receiver allow himself to be fooled by an easily faked, approximately cheap talk message from an *enemy*?

Puzzles

In all three examples (two of them nearly real), the message-sender deceived the message-receiver and won, but in the *less* beneficial of the two possible ways

- Why did the receiver allow himself to be fooled by an easily faked, approximately cheap talk message from an *enemy*?

- And if the sender expected his deception to succeed, why didn't he reverse the message and win in the more beneficial way?

  (Why didn’t Alpha say he’d be dining on Saturday and book dinner Sunday?
  Why didn't the Allies feint at Normandy and attack at Calais?
  Why didn't Kongming light fires and ambush Cao Cao on the Main Road?)

An analysis should also reconcile the answers to these puzzles with the surface differences in the senders' messaging strategies and the receivers' responses:

- Alpha's and the Allies' messages were literally lies, but they were sent with the belief that Gamma and the Germans would believe them or—perhaps expecting a double bluff—invert them one too many times

- Kongming's message was literally true, but Cao Cao, expecting a lie, was fooled by Kongming's double bluff into inverting the message

An analysis should also reconcile the answers to these puzzles with the surface differences in the senders' messaging strategies and the receivers' responses:

- Alpha's and the Allies' messages were literally lies, but they were sent with the belief that Gamma and the Germans would believe them or—perhaps expecting a double bluff—invert them one too many times

- Kongming's message was literally true, but Cao Cao, expecting a lie, was fooled by Kongming's double bluff into inverting the message

In this case Luo Guanzhong actually tells us what the generals were thinking:

- Kongming: "Have you forgotten the tactic of 'letting weak points look weak and strong points look strong'?"

- Cao Cao: "Don't you know what the military texts say? 'A show of force is best where you are weak. Where strong, feign weakness.'"

An analysis should also reconcile the answers to these puzzles with the surface differences in the senders' messaging strategies and the receivers' responses:

- Alpha's and the Allies' messages were literally lies, but they were sent with the belief that Gamma and the Germans would believe them or—perhaps expecting a double bluff—invert them one too many times

- Kongming's message was literally true, but Cao Cao, expecting a lie, was fooled by Kongming's double bluff into inverting the message

In this case Luo Guanzhong actually tells us what the generals were thinking:

- Kongming: "Have you forgotten the tactic of 'letting weak points look weak and strong points look strong'?"

- Cao Cao: "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

Puffery: Tale of Two Henrys

Henry Chichele

Henry VI

Henry Chichele, sometime Archbishop of Canterbury, his mind having been stirred up by his Almighty Father, wishes to persuade his King, Henry VI, to endow a college with liberal maintenance for His servants

Both Henrys would like the faithful departed in Oxfordshire to be released from Purgatory as soon as possible

But only Chichele knows how much the liberality of the college's maintenance will accelerate their release

Henry Chichele, sometime Archbishop of Canterbury, his mind having been stirred up by his Almighty Father, wishes to persuade his King, Henry VI, to endow a college with liberal maintenance for His servants

Both Henrys would like the faithful departed in Oxfordshire to be released from Purgatory as soon as possible

But only Chichele knows how much the liberality of the college's maintenance will accelerate their release

And Chichele, unlike his King, has an additional, temporal reason for wanting the maintenance to be as liberal as possible

Henry Chichele, sometime Archbishop of Canterbury, his mind having been stirred up by his Almighty Father, wishes to persuade his King, Henry VI, to endow a college with liberal maintenance for His servants

Both Henrys would like the faithful departed in Oxfordshire to be released from Purgatory as soon as possible

But only Chichele knows how much the liberality of the college's maintenance will accelerate their release

And Chichele, unlike his King, has an additional, temporal reason for wanting the maintenance to be as liberal as possible (sinecures for pesky nephews)

Henry Chichele, sometime Archbishop of Canterbury, his mind having been stirred up by his Almighty Father, wishes to persuade his King, Henry VI, to endow a college with liberal maintenance for His servants

Both Henrys would like the faithful departed in Oxfordshire to be released from Purgatory as soon as possible

But only Chichele knows how much the liberality of the college's maintenance will accelerate their release

And Chichele, unlike his King, has an additional, temporal reason for wanting the maintenance to be as liberal as possible (sinecures for pesky nephews)

In 1436, Chichele sends a nuanced proposal to Henry VI: "Sire, I recommend that you endow this college with really, really, really… [x times] liberal maintenance"

Henry VI counts the "really"s, discounts for possible adverb inflation, and chooses how liberal to make the new college's maintenance

Puffery: Blodget (Wang, Spezio, and Camerer, 2010 *American Economic Review*, "WSC"; based on Crawford-Sobel, 1982 *Econometrica*). WSC's frame:

> During the tech-stock bubble, Wall Street security analysts were alleged to inflate recommendations about the future earnings prospects of firms in order to win investment banking relationships with those firms. Specifically, analysts of Merrill Lynch used a five-point rating system (1 = Buy to 5 = Sell) to predict how the stock would perform. They usually gave two 1–5 ratings for short run (0–12 months) and long run (more than 12 months) performance separately.
>
> Henry Blodget, Merrill Lynch's famously optimistic analyst, "did not rate any Internet stock a 4 or 5" during the bubble period (1999 to 2001). In one case, the online direct marketing firm LifeMinders, Inc. (LFMN), Blodget first reported a rating of 2-1 (short run "accumulate"—long run "buy") when Merrill Lynch was pursuing an investment banking relationship with LFMN. Then, the stock price gradually fell from $22.69 to the $3–$5 range. While publicly maintaining his initial 2-1 rating, Blodget privately e-mailed fellow analysts that "LFMN is at $4. I can't believe what a POS [piece of shit] that thing is." He was later banned from the security industry for life and fined millions of dollars.
>
> (Source: Complaint, Order, and Final Judgement in Securities and Exchange Commission v. Henry M. Blodget, (2003) Civ. 2947 (WHP) (S.D.N.Y.).)

Blodget, like Two Henrys, is a sender-receiver game between the analyst and a single investor (the firm whose stock is being touted plays no active role):

- The analyst has private information about the stock's prospects
- The analyst's recommendation to the investor is approximately cheap talk
- Based on that recommendation, the investor makes a decision that affects the analyst's welfare as well as the investor's own welfare
- The analyst's and investor's preferences are similar, in that both want the investor to sell on bad news and buy/hold on good, other things equal
- But there is a wedge between their preferences, in that the analyst's desire to preserve his relationship with the firm makes him want the investor to buy/hold the stock more than a well-informed investor would

Puzzle: Why would anyone be fooled by a cheap talk message from someone whose interests are different?

Gullibility: Gibbous Grass?

[From the College betting book; my scholia in square brackets]

[Warden] Davis bets [Fellow whose study overlooked the Gt Quad] Ryan 1 doz oysters that the lawn in the Gt Quad is more circular than it is gibbous or oval.

Adjudicant[s]: Perkins, Häcker

[Signed]: J Davis MJ Ryan
1.X.05 [a Saturday]

[To paraphrase Potter Stewart, the bettors and adjudicants seem to have thought they would know eccentricity when they saw it]

Gullibility: Gibbous Grass?

[From the College betting book; my scholia in square brackets]

[Warden] Davis bets [Fellow whose study overlooked the Gt Quad] Ryan 1 doz oysters that the lawn in the Gt Quad is more circular than it is gibbous or oval.

Adjudicant[s]: Perkins, Häcker

[Signed]: J Davis MJ Ryan
1.X.05 [a Saturday]

[To paraphrase Potter Stewart, the bettors and adjudicants seem to have thought they would know eccentricity when they saw it…and so they did]

Ryan wins
BH
J Perkins

Gullibility: Cider in your Ear?

"Son…One of these days in your travels, a guy is going to show you a brand-new deck of cards on which the seal is not yet broken. Then this guy is going to offer to bet you that he can make the jack of spades jump out of this brand-new deck of cards and squirt cider in your ear. But, son, do not accept this bet, because as sure as you stand there, you're going to wind up with an ear full of cider."

—Obadiah ("The Sky") Masterson, quoting his father in Damon Runyon (*Guys and Dolls: The Stories of Damon Runyon*, 1932)

Gullibility: Cider in your Ear?

> "Son…One of these days in your travels, a guy is going to show you a brand-new deck of cards on which the seal is not yet broken. Then this guy is going to offer to bet you that he can make the jack of spades jump out of this brand-new deck of cards and squirt cider in your ear. But, son, do not accept this bet, because as sure as you stand there, you're going to wind up with an ear full of cider."
>
> —Obadiah ("The Sky") Masterson, quoting his father in Damon Runyon (*Guys and Dolls: The Stories of Damon Runyon*, 1932)

To economic theorists, this quotation brings to mind Milgrom and Stokey's, 1982 *Journal of Economic Theory*, celebrated "No Trade Theorem" (illustrated below)

The No Trade Theorem was later dubbed the "Groucho Marx Theorem"

> "I sent the club a wire stating, 'Please accept my resignation. I don't want to belong to any club that will accept people like me as a member'."
>
> —Groucho Marx, Telegram to the Beverly Hills Friars' Club

Can traditional game theory resolve these puzzles? Nash equilibrium

- In a game, two or more players choose strategies, which jointly determine their welfares or payoffs; any uncertainty is handled by assuming that players' welfares are represented by expected payoffs

- A Nash equilibrium is a profile of players' strategies in which each player's strategy maximizes his expected payoff, given the others' strategies

- A Nash equilibrium is thus a kind of rational expectations equilibrium, in which players form expectations or beliefs about each other's strategies that are self-confirming if players choose best responses to their beliefs

- Refinements like subgame-perfect equilibrium extend this notion to games with sequential decisions; refinements like Bayesian, perfect Bayesian, or sequential equilibrium extend it to games with asymmetric information

- Either way, equilibrium builds in rationality of individual decisions, but bundles it with the far stronger assumption that players' beliefs are coordinated

Equilibrium in Animarumadversion, D-Day, and Huarongdao (Crawford-Sobel, 1982 *Econometrica*; Crawford, 2003 *American Economic Review*)

- In each case one player, the sender, sends a nonbinding, cheap talk message to the receiver regarding his planned action; lying has no direct cost
- The receiver observes the message
- The sender and receiver make decisions that jointly determine their payoffs
- The sender's and receiver's payoffs are (at least approximately) opposed

Equilibrium in Animarumadversion, D-Day, and Huarongdao (Crawford-Sobel, 1982 *Econometrica*; Crawford, 2003 *American Economic Review*)

- In each case one player, the sender, sends a nonbinding, cheap talk message to the receiver regarding his planned action; lying has no direct cost

- The receiver observes the message

- The sender and receiver make decisions that jointly determine their payoffs

- The sender's and receiver's payoffs are (at least approximately) opposed

In games like these, in any equilibrium the sender's cheap talk message must be uninformative, and the receiver must ignore it

For, if the sender made his message informative, the receiver's optimal response to it would increase the receiver's own payoff and thus reduce the sender's, who would therefore do better by making his message uninformative

Equilibrium behavior therefore makes the communication phase irrelevant

Yet in the world, deception is common and succeeds, even in zero-sum games

Equilibrium in Two Henrys and Blodget (Crawford-Sobel, 1982 *Econometrica*)

- In each case one player, the sender, observes a signal relevant to both their payoffs and sends a cheap talk message about it; lying has no direct cost
- The receiver observes the message and makes a decision that, with the sender's true signal, determines both the sender's and the receiver's payoffs
- The sender's and receiver's preferences about how the receiver's decision relates to the sender's signal are qualitatively similar, but differ systematically

Equilibrium in Two Henrys and Blodget (Crawford-Sobel, 1982 *Econometrica*)

- In each case one player, the sender, observes a signal relevant to both their payoffs and sends a cheap talk message about it; lying has no direct cost
- The receiver observes the message and makes a decision that, with the sender's true signal, determines both the sender's and the receiver's payoffs
- The sender's and receiver's preferences about how the receiver's decision relates to the sender's signal are qualitatively similar, but differ systematically

Crawford and Sobel showed that in all equilibria, the sender divides the possible signals into intervals and tells the receiver only which interval the signal fell in

There is always a "babbling" equilibrium, and when the sender's and receiver's preferences are sufficiently far apart it is unique, except for messaging variations

Equilibrium in Two Henrys and Blodget (Crawford-Sobel, 1982 *Econometrica*)

- In each case one player, the sender, observes a signal relevant to both their payoffs and sends a cheap talk message about it; lying has no direct cost
- The receiver observes the message and makes a decision that, with the sender's true signal, determines both the sender's and the receiver's payoffs
- The sender's and receiver's preferences about how the receiver's decision relates to the sender's signal are qualitatively similar, but differ systematically

Crawford and Sobel showed that in all equilibria, the sender divides the possible signals into intervals and tells the receiver only which interval the signal fell in

There is always a "babbling" equilibrium, and when the sender's and receiver's preferences are sufficiently far apart it is unique, except for messaging variations

There are also more informative equilibria when the sender's and receiver's preferences are closer together, but all have intentional vagueness

With closer preferences, more information can be transmitted in equilibrium

This is not completely unhelpful; but equilibrium/rational expectations doesn't explain systematic deception, or why senders tend to lie in the direction that would push credulous receivers in their favored direction, or why senders are more truthful and receivers more credulous than any equilibrium predicts

Equilibrium in Gibbous Grass? and Cider in your Ear?
(Brocas, Carillo, Camerer, and Wang, 2014 *Review of Economic Studies*)

Brocas et al. ran experiments on simple three-state betting games (close to zero-sum; game-theoretic analogues of Milgrom and Stokey's market trading model)

There are three ex ante equally likely states, A, B, C

Player 1 learns privately either that the state is {A or B} or that it is C

Player 2 learns privately either that the state is A or that it is {B or C}

Player/state	**A**	**B**	**C**
1	25	5	20
2	0	30	5

Players then choose simultaneously whether to Bet or Pass

A player who chooses Pass, or who chooses Bet while the other player chooses Pass, earns 10, whatever the state

If both choose Bet, they get their payoffs in the table for whichever state occurs

All this is publicly announced (to induce common knowledge)

This game has a unique sensible equilibrium, identifiable via 3 rounds of “iterated weak dominance” (there’s a nonsensical equilibrium in which both always Pass)

Round 1 of iterated weak dominance (Bet, Pass)

player/state	A	B	C
1	25	5	20
2	0	30	5

Round 2

player/state	A	B	C
1	25	5	20
2	0	30	5

Round 3

player/state	A	B	C
1	25	5	20
2	0	30	5

In equilibrium, there is no betting in any state (player 1 is *willing* to bet in state C)

But real people make zero-sum bets all the time, in predictable patterns

Can behavioral game theory do better?

With enough stationary repetition, even amoebas can learn to play an equilibrium

But when people interact in a new setting, equilibrium requires strategic thinking

Equilibrium thinking often involves complex fixed-point or indefinitely iterated rationality-based reasoning, which even quants find unnatural or inaccessible

Humans may then find simpler, nonequilibrium ways of thinking about the game

Can behavioral game theory do better?

With enough stationary repetition, even amoebas can learn to play an equilibrium

But when people interact in a new setting, equilibrium requires strategic thinking

Equilibrium thinking often involves complex fixed-point or indefinitely iterated rationality-based reasoning, which even quants find unnatural or inaccessible

Humans may then find simpler, nonequilibrium ways of thinking about the game

Yet even those who grant the desirability of modeling strategic thinking more realistically, have long doubted whether that is feasible:

- How can any model systematically out-predict a rational-expectations notion?
- And how can one hope to identify such a model among the plethora of logically possible non-equilibrium models of strategic thinking?

Level-*k* models of strategic thinking

These questions are answered, to some extent, by a body of experimental work that studies strategic thinking by eliciting initial responses to games (surveyed in Crawford, Costa-Gomes, and Iriberri, 2013 *Journal of Economic Literature*).

In all but the simplest games, subjects' thinking avoids the fixed-point or indefinitely iterated dominance reasoning that equilibrium usually requires

Level-*k* models of strategic thinking

These questions are answered, to some extent, by a body of experimental work that studies strategic thinking by eliciting initial responses to games (surveyed in Crawford, Costa-Gomes, and Iriberri, 2013 *Journal of Economic Literature*).

In all but the simplest games, subjects' thinking avoids the fixed-point or indefinitely iterated dominance reasoning that equilibrium usually requires

Instead subjects tend to follow "level-*k*" rules of thumb that:

- anchor their beliefs in a naïve model of others' decisions, called *L0*, and
- adjust their beliefs via a small, heterogeneous number (*k*) of iterated best responses: *L1* best responds to *L0*, *L2* to *L1*, and so on

Level-*k* models of strategic thinking

These questions are answered, to some extent, by a body of experimental work that studies strategic thinking by eliciting initial responses to games (surveyed in Crawford, Costa-Gomes, and Iriberri, 2013 *Journal of Economic Literature*).

In all but the simplest games, subjects' thinking avoids the fixed-point or indefinitely iterated dominance reasoning that equilibrium usually requires

Instead subjects tend to follow "level-*k*" rules of thumb that:

- anchor their beliefs in a naïve model of others' decisions, called *L0*, and
- adjust their beliefs via a small, heterogeneous number (*k*) of iterated best responses: *L1* best responds to *L0*, *L2* to *L1*, and so on

- *Lk* (for $k > 0$) is decision-theoretically rational, with an accurate model of the game; it differs from equilibrium only in deriving beliefs from a simpler rule

- *Lk* mimics equilibrium decisions in simple (e.g. *k*-dominance-solvable) games, but may deviate systematically in more complex games; its deviations make it possible for it to out-predict a rational-expectations notion such as equilibrium

Compare Keynes’s 1936 *General Theory* comparison of 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. 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.* [emphasis added]

Compare Keynes's 1936 *General Theory* comparison of 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. 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.* [emphasis added]

[Keynes could read some German, but there's no sign he had read von Neumann's 1928 *Mathematische Annalen* paper (with the first equilibrium existence theorem), so he may have been less tempted to look for a notion like equilibrium than we are]

Level-*k* thinking in Animarumadversion, D-Day, and Huarongdao
(Crawford, 2003 *American Economic Review*)

The evidence from games without communication is largely consistent with an *L0* that is uniform random over the feasible decisions

Such a "random" *L0* seems to reflect higher levels' thinking about the incentives the payoff structure creates, before they begin to consider others' incentives

In games with communication, however, intuition and the (limited) evidence suggest that the first thing we do when hearing a message, even from an enemy, is to try to understand its literal meaning, before we consider others' incentives

This motivates anchoring *L0* in truthfulness for senders or credulity for receivers

Higher levels are defined by iterated best responses as before

(The Luo Guanzhong quotations suggest Kongming was *L3* and Cao Cao *L2*)

My 2003 analysis assumed that the sender and receiver are each drawn from a population including both level-*k* and *Sophisticated* players

- Level-*k* players avoid fixed-point reasoning, anchor beliefs on truthfulness or credulity, and determine beliefs and decisions by iterated best responses

- *Sophisticated* players choose equilibrium decisions in a reduced game that reflects the possibility and frequencies of level-*k* and *Sophisticated* players

My 2003 analysis assumed that the sender and receiver are each drawn from a population including both level-*k* and *Sophisticated* players

- Level-*k* players avoid fixed-point reasoning, anchor beliefs on truthfulness or credulity, and determine beliefs and decisions by iterated best responses

- *Sophisticated* players choose equilibrium decisions in a reduced game that reflects the possibility and frequencies of level-*k* and *Sophisticated* players

The main goal was to learn whether and when the possibility of level-*k* players in each role allows *Sophisticated* senders to “deceive” *Sophisticated* receivers

My 2003 analysis assumed that the sender and receiver are each drawn from a population including both level-*k* and *Sophisticated* players

- Level-*k* players avoid fixed-point reasoning, anchor beliefs on truthfulness or credulity, and determine beliefs and decisions by iterated best responses

- *Sophisticated* players choose equilibrium decisions in a reduced game that reflects the possibility and frequencies of level-*k* and *Sophisticated* players

The main goal was to learn whether and when the possibility of level-*k* players in each role allows *Sophisticated* senders to "deceive" *Sophisticated* receivers

That some players might be level-*k* completely alters the game's character

- Because *Sophisticated* players' payoffs are influenced by level-*k* players' decisions, the game is no longer zero-sum and messages no longer cheap talk

- Unlike the underlying game, the reduced game has asymmetric information about players' behavioral rules; a *Sophisticated* receiver reads the sender's message about his intentions as an informative signal of the sender's rule

- If *Sophisticated* senders and receivers have high population frequencies, the reduced game has a unique mixed-strategy equilibrium, which is outcome-equivalent to the equilibrium of the game without communication

- If *Sophisticated* senders and receivers have low frequencies, the reduced game has an essentially unique pure-strategy equilibrium, in which *Sophisticated* senders send the message that deceives the most frequent kind of level-*k* receiver, and then try for the *less* beneficial way to win

- If *Sophisticated* senders and receivers have high population frequencies, the reduced game has a unique mixed-strategy equilibrium, which is outcome-equivalent to the equilibrium of the game without communication

- If *Sophisticated* senders and receivers have low frequencies, the reduced game has an essentially unique pure-strategy equilibrium, in which *Sophisticated* senders send the message that deceives the most frequent kind of level-*k* receiver, and then try for the *less* beneficial way to win

- In the latter case (as in the former), there is never a sensible equilibrium in which *Sophisticated* senders try for the *more* beneficial way to win

  (For, in such an equilibrium any deviation from the *Sophisticated* sender's equilibrium message would "prove" to a *Sophisticated* receiver that the sender is level-*k*, leading a *Sophisticated* receiver to try for the less beneficial way to win, and thus leading a *Sophisticated* sender to try for the less beneficial way)

- Thus, with no unexplained difference in the *Sophistication* of senders and Receivers, and for plausible parameter values, the level-*k* model explains why *Sophisticated* receivers might allow themselves to be "deceived", and why *Sophisticated* senders don't try for the more beneficial way to win

Level-*k* thinking in Two Henrys and Blodget
(Crawford-Sobel, 1982 *Econometrica*; Crawford, 2003 *American Economic Review*; Kartik, Ottaviani, and Squintani, 2007 *Journal of Economic Theory*; Wang, Spezio, and Camerer, 2010 *American Economic Review*)

WSC's experimental design closely follows their Blodget example:

- A sender observes the state, S = 1, 2, 3, 4, or 5, and sends a message, M = 1, 2, 3, 4, or 5 (The clear correspondence between state and message labelings ensures that messages are understood, and makes lying meaningful)

- A receiver observes the message M, and chooses an action A = 1, 2, 3, 4, or 5, which together with S determines his and the sender's welfare

- Senders and receivers have single-peaked preferences, with the receiver's ideal outcome A = S and the sender's A = S + b (ignoring boundaries)

- The design varies the difference between the sender's and the receiver's preferences across three treatments: b = 0, 1, or 2

WSC focused on the most informative equilibria in their games, as benchmarks

In WSC's Figures 1-3, copied below, a circle's size shows senders' message frequencies (columns) in the various states (rows) and a circle's darkness and the numbers inside it show receivers' action frequencies

- In Figure 1 the sender's and receiver's preferences are identical (b = 0); the most informative equilibrium has truth-telling and credulity: M = S and A = S

  There are no significant deviations from that equilibrium

WSC focused on the most informative equilibria in their games, as benchmarks

In WSC's Figures 1-3, copied below, a circle's size shows senders' message frequencies (columns) in the various states (rows) and a circle's darkness and the numbers inside it show receivers' action frequencies

- In Figure 1 the sender's and receiver's preferences are identical (b = 0); the most informative equilibrium has truth-telling and credulity: M = S and A = S

  There are no significant deviations from that equilibrium

- In Figure 2 the sender's and receiver's preferences differ somewhat (b = 1); the most informative equilibrium has the sender sending M = 1 when S = 1 and the receiver responding with A = 1; and otherwise the sender's message distribution is the same for S = 2, 3, 4, 5, and the receiver responds A = 3 or 4

  Both senders and receivers deviate systematically from that equilibrium

  Senders lie in the direction (above the diagonal) that would make credulous receivers choose actions senders would prefer, while making messages more truthful than in the equilibrium (M distributions shift right as S goes from 2 to 5) And receivers are more credulous (A > S, A > best response to senders)

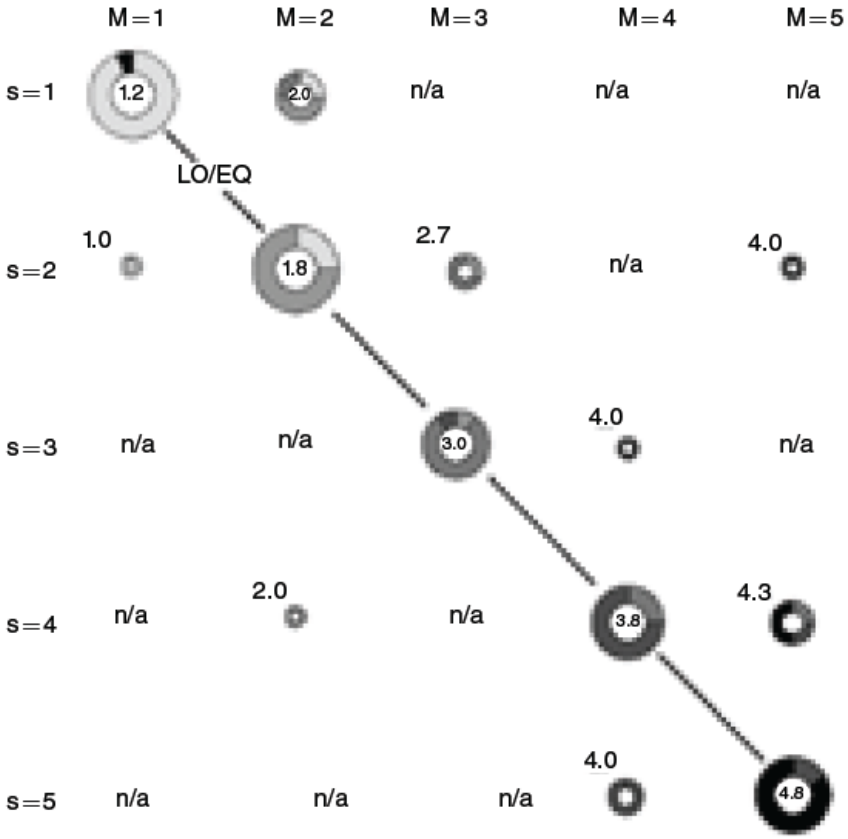

FIGURE 1. RAW DATA PIE CHARTS ($b = 0$)
(HIDDEN BIAS–STRANGER)

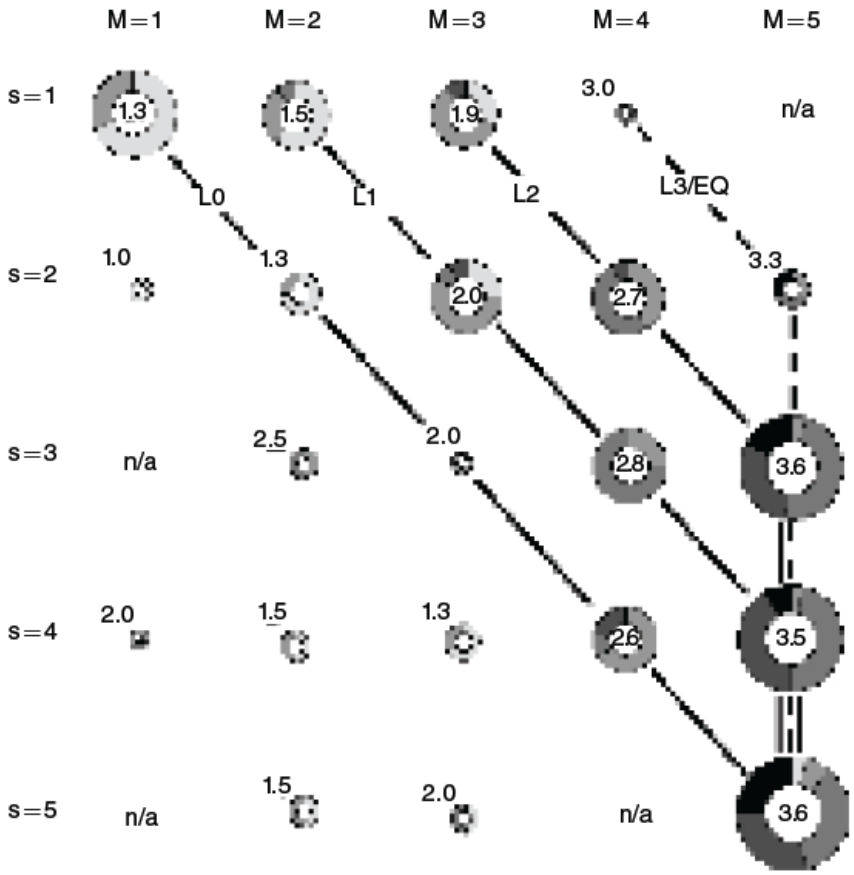

FIGURE 2. RAW DATA PIE CHART ($b = 1$)
(HIDDEN BIAS–STRANGER)

- In Figure 3 the sender's and receiver's preferences differ a great deal (b = 2); the only equilibria are babbling equilibria, in which the sender's message distribution is the same for all S, and the receiver ignores the sender's messages and chooses A = 3, the optimal action given the receiver's prior

  Both senders and receivers deviate systematically from that equilibrium

  Senders again lie in the direction (above diagonal) that would make credulous receivers choose actions sender would prefer, while making messages more truthful than in equilibrium (M distributions shift right as S goes from 1 to 5)

  Receivers are again more credulous (A > S, > best response to senders)

- Despite the systematic deviations from equilibrium when b = 1 or 2, the amount of information transmitted, measured by the correlation between S and A, declines with the distance between the sender's and receiver's preferences, as suggested by Crawford-Sobel's equilibrium-based comparative statics result

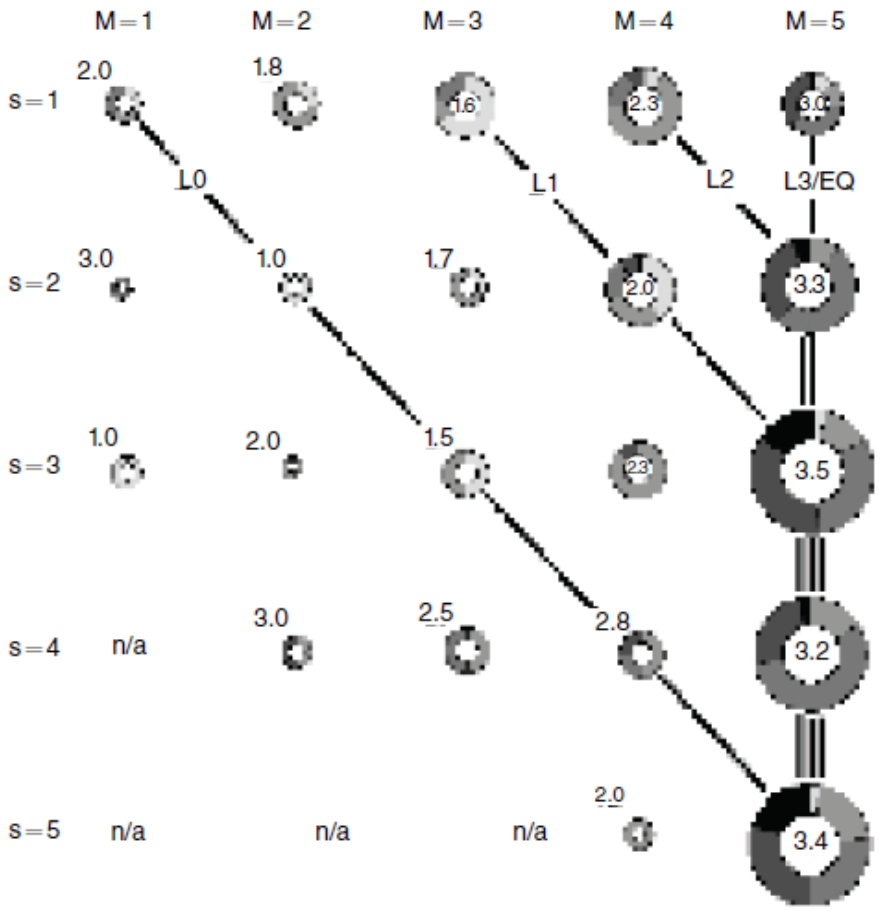

FIGURE 3. RAW DATA PIE CHART ($b = 2$)
(HIDDEN BIAS–STRANGER)

WSC used a level-*k* model like Crawford's, 2003 *American Economic Review* model and Kartik, Ottaviani, and Squintani's, 2007 *Journal of Economic Theory* model to analyze their results, again assuming lying has no direct cost

A receiver's best outcome is A = S; ignoring boundaries, a sender's is A = S + b

- In the level-*k* model, players anchor beliefs in a truthful sender *L0*, which sets M = S; and a credulous receiver *L0*, which sets A = M

- *L1* senders best respond to *L0* receivers, "puffing" their messages by b: M = S + b (ignoring boundaries, here and below), so *L0* receivers choose S + b, which would yield an *L1* sender's best action, given a credulous receiver

- *L1* receivers best respond to *L1* senders, de-puffing messages by b: A = M – b, which would yield an *L1* receiver's ideal action, given her/his belief that *L1* senders best respond to *L0* receivers, setting M = S + b

- *L2* senders best respond to *L1* receivers, puffing by 2d: M = S + 2b; *L2* receivers best respond to *L2* senders, de-puffing by 2b: A = M – 2b; and so on

The labels in Figures 1-3 show a close association between senders' and receivers' decisions and *L1*, *L2*, or *L3* behavior

(Explicitly labeled in Figures 2 and 3; same as equilibrium behavior in Figure 1)

Overall, the level-*k* model gives a unified explanation of the main fact patterns:

- Senders lie in the direction that would make credulous receivers choose actions the Sender would prefer, trying to outguess receivers' discounting

- Senders' messages are nonetheless more truthful than in any equilibrium

- Receivers are more credulous than in any equilibrium

The labels in Figures 1-3 show a close association between senders' and receivers' decisions and *L1*, *L2*, or *L3* behavior

(Explicitly labeled in Figures 2 and 3; same as equilibrium behavior in Figure 1)

Overall, the level-*k* model gives a unified explanation of the main fact patterns:

- Senders lie in the direction that would make credulous receivers choose actions the Sender would prefer, trying to outguess receivers' discounting

- Senders' messages are nonetheless more truthful than in any equilibrium

- Receivers are more credulous than in any equilibrium

Even though the model makes lying costless, *Lk* behavior is anchored in a truthful or credulous *L0*; this gives *Lk* a residue of truthfulness or credulity that only equilibrium reasoning would completely massage away

The sensitivity of *Lk*'s behavior to the distance between sender's and receiver's preferences also explains why Crawford-Sobel's equilibrium comparative statics result is qualitatively robust to large, systematic deviations from equilibrium

Level-*k* thinking in Gibbous Grass? and Cider in your Ear?
(Crawford and Iriberri, 2007 *Econometrica*; Brocas, Carillo, Camerer, Wang, 2014 *Review of Economic Studies*)

- In games with asymmetric information, I take *L0*'s decisions to be uniform over the feasible decisions, and *independent of its own value*

  (This may seem odd, but *L0* is not an actual player: It is a player's naïve model of other players whose values he does not observe; reasoning contingent on others' possible values is logically possible, but behaviorally far-fetched)

- Higher levels are defined by iterated best responses as before

Level-*k* thinking in Gibbous Grass? and Cider in your Ear?
(Crawford and Iriberri, 2007 *Econometrica*; Brocas, Carillo, Camerer, Wang, 2014 *Review of Economic Studies*)

- In games with asymmetric information, I take *L0*'s decisions to be uniform over the feasible decisions, and *independent of its own value*

  (This may seem odd, but *L0* is not an actual player: It is a player's naïve model of other players whose values he does not observe; reasoning contingent on others' possible values is logically possible, but behaviorally far-fetched)

- Higher levels are defined by iterated best responses as before

This "random" level-*k* model gives a realistic account of people's "informational naiveté", failure to attend to how others' incentives depend on their information

- Sky Masterson's father was worried that his son would be an *L1* defined this way: rational but insufficiently skeptical of offers that are "too good to be true"

- Milgrom and Stokey, speculating on why zero-sum trades occur despite their Groucho Marx Theorem, conjecture the rules Naïve Behavior, which sticks with its prior but otherwise behaves rationally, like this model's *L1*; and First-Order Sophistication, which best responds to Naïve Behavior, like this model's *L2*

In a careful clustering analysis that lets the data speak directly, Brocas et al. use the random level-*k* model to interpret experimental results on zero-sum betting

Recall that equilibrium predicts no betting, in any state

player/state	A	B	C
1	25	5	20
2	0	30	5

Yet, as in several similar previous experiments, half of Brocas et al.'s subjects Bet, in patterns that varied systematically with the player role and state

The level-*k* model makes specific predictions of betting patterns

L1 respects only simple dominance:

player/state	A	B	C
1	25	5	20
2	0	30	5

L2 respects two rounds of iterated weak dominance:

player/state	A	B	C
1	25	5	20
2	0	30	5

And *L3* respects three rounds (enough for equilibrium in this game):

player/state	A	B	C
1	25	5	20
2	0	30	5

If all subjects were *L1s*, 100% of player 1s and 67% of player 2s would be willing to bet, with 100% betting in states B and C, each many more than in the data

player/state	A	B	C
1	25	5	20
2	0	30	5

But Brocas et al. find clusters of subjects whose behavior corresponds to each of *L1*, *L2*, and *L3*; and also a cluster of "irrational" players

*L2*s and *L3*s are less gullible than *L1*s, and the level-*k* model with estimated level frequencies fits better than equilibrium or any homogeneous model

Interesting roads not taken here

Sobel, "A Theory of Credibility," 1985 *REStud*, studies repeated interactions with cheap talk messages, relaxing the assumption that Receivers know the Sender's motives. Friendly Senders always tell the truth; enemy Senders tell the truth until there is a sufficiently important opportunity to lie, and then cash in their reputation

Matthew Gentzkow https://gentzkow.people.stanford.edu/ with various co-authors (e.g. Allcott and Gentzkow, "Social Media and Fake News in the 2016 Election," 2017 *JEP*) and Philipp Howard http://philhoward.org/ with various co-authors (e.g. Howard, Woolley, and Calo, "Algorithms, Bots, and Political Communication in the US 2016 Election: The Challenge of Automated Political Communication for Election Law and Administration," 2018 *J. Information Technology and Politics*) study various aspects of the political economy of news and social media, with particular attention to the consumption value of news and deceptive news

Dixit and Weibull, "Political Polarization," 2007 *PNAS*, and Roux and Sobel, "Group Polarization in a Model of Information Aggregation," 2015 *AEJ: Micro* (though not strictly about deception or gullibility) study settings in which groups of rational Bayesian individuals with differing priors update their beliefs based on the same information, yet still sometimes become more polarized