Behavioural Economics:
Strategic Thinking

University of Oxford, Michaelmas Term 2014

**Vincent P. Crawford, University of Oxford, All Souls College,
and University of California, San Diego**

**(with thanks to Miguel Costa-Gomes, Nagore Iriberri,
Colin Camerer, Teck-Hua Ho, and Juin-Kuan Chong)**

Revised 16 June 2014

Strategic Thinking

Strategic thinking pervades human interaction.

As soon as children develop enough “theory of mind” to model other people as independent decision makers, they must be *taught* to look both ways before crossing one-way streets—suggesting that they instinctively rely on rationality assumptions to predict others’ decisions.

Adult attempts to predict others’ responses to incentives are shaped by similar—though usually more subtle—rationality-based inferences.

Yet from a behavioural/empirical point of view, strategic thinking has been downplayed in economics and game theory.

The canonical model of strategic thinking is the game-theoretic notion of Nash equilibrium, defined as a combination of strategies, one for each player, such that each player's strategy maximizes his expected payoff, given the others' strategies.

(Equilibrium can be defined and applied without reference to its interpretation, but it is best thought of as an "equilibrium in beliefs," in which players' equilibrium strategies represent beliefs about others' strategies that are correct given the rational strategy choices they imply.)

Nash equilibrium addresses the problem that in games, decision-theoretic rationality alone seldom restricts behavior enough to be useful.

Even common knowledge of rationality implies only that players' strategies are rationalizable (Bernheim 1984 *Econometrica* and Pearce 1984 *Econometrica*), which leaves behavior unrestricted in many games.

Equilibrium makes more definite predictions by augmenting rationality with the "rational-expectations" assumption that players' beliefs are correct, and therefore the same for all players.

Because many games have multiple equilibria, equilibrium is often augmented by refinements, with the goal of deriving unique predictions.

Aside on common or finitely iterated knowledge of rationality

In two-person games (with some differences in n-person games that are unimportant here), the implications of common or finitely iterated knowledge of players' rationality (without further restrictions on beliefs) are captured by finitely iterated strict dominance and *k*-rationalizability.

k-rationalizability reflects the implications of *k* levels of mutual knowledge of rationality; rationalizability is equivalent to *k*-rationalizability for all *k*.

A 1-rationalizable strategy (the sets R1 on the next slide) is one for which there is a profile of others' strategies that makes it a best response.

A 2-rationalizable strategy (the sets R2) is one for which there exists a profile of others' 1-rationalizable strategies that make it a best response.

And so on….

Each generally yields set-valued restrictions on individual players' strategy choices (unlike equilibrium which restricts their relationship).

Each game has a unique equilibrium (M, C). In the first game M and C are the only rationalizable strategies; in the second game all strategies are rationalizable, for each player.

		R1,R2	R1,R2,R3,R4	
		L	C	R
R1,R2,R3	T	0 7	5 0	3 0
R1,R2,R3,R4	M	0 5	2 2	0 5
R1	B	7 0	5 0	3 7

Dominance-solvable game

		Rk for all k	Rk for all k	Rk for all k
		L	C	R
Rk for all k	T	0 7	5 0	7 0
Rk for all k	M	0 5	2 2	0 5
Rk for all k	B	7 0	5 0	0 7

Unique equilibrium without dominance

Equilibrium reflects the implications of common knowledge of rationality *plus* common beliefs:

Any equilibrium strategy is *k*-rationalizable for all *k*, but not all combinations of rationalizable strategies are in equilibrium.

In games that are dominance-solvable in *k* rounds, *k*-rationalizability implies that players have the same beliefs—with a qualification for mixed-strategy equilibria that is unimportant here—so any combination of *k*-rationalizable strategies is in equilibrium as in the first game above.

In other games, *k*-rationalizability and rationalizability allow deviations from equilibrium as in the second game above, where there is a “tower” or “helix” of beliefs, consistent even with common knowledge of rationality, to support any combination of strategies.

(But except for the equilibrium beliefs (M, C), the beliefs in the tower or helix differ across players, and many are behaviourally implausible.)

End of aside

The generality, simplicity, and tractability of equilibrium analysis have made it the method of choice in strategic applications.

If a setting allows learning, and if only long-run outcomes matter, and if equilibrium is unique or equilibrium selection does not depend on the details of learning, then applications can safely assume equilibrium.

But otherwise, if equilibrium is justified, it must be via strategic thinking.

Epistemic game theory provides conditions under which reasoning based on iterated knowledge of rationality and beliefs focuses players' beliefs on a particular equilibrium, even in their initial responses to a game.

But in many games such reasoning is complex enough to make the thinking justification for equilibrium behaviorally implausible.

Even people who are capable of such thinking may doubt that others are capable, and therefore be unwilling to play their part of an equilibrium.

In such settings, reliably predicting initial responses to games may require a non-equilibrium model of strategic thinking.

Modeling strategic thinking more accurately can yield several benefits:

- It can establish the robustness of conclusions based on equilibrium in games where empirically reliable rules mimic equilibrium

- It can also challenge conclusions based on equilibrium or refinements in games where equilibrium is implausible without learning

- It can resolve empirical puzzles by explaining the deviations from equilibrium that some games evoke

- It can also elucidate the structure of learning from imperfect analogies, where assumptions about cognition determine which analogies between current and previous games players recognize and elucidate the structure of learning, distinguishing reinforcement from beliefs-based and more sophisticated rules.

However, even those who grant the desirability of improving upon equilibrium models of initial responses may doubt its feasibility.

How can any model systematically out-predict a rational-expectations notion such as equilibrium?

And how can one identify simple models that allow such improvements among the huge variety of logically possible non-equilibrium models?

Experimental research shows with increasing clarity that people's initial responses to games often deviate systematically from equilibrium.

Importantly, the results also show that the deviations have a large structural component that can be modeled in a simple, tractable way.

This component exists because subjects' thinking tends to avoid the fixed-point or indefinitely iterated dominance reasoning that equilibrium sometimes requires. In Selten's words:

> "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."

(This is not to say that with enough experience in a stationary setting, learning can't make people converge to steady states that *we* would need fixed-point reasoning to characterize, just that such reasoning doesn't directly describe people's thinking.)

If subjects' thinking avoids fixed-point or indefinitely iterated dominance reasoning, then what does it consist of?

The experimental evidence (see especially Nagel 1995 *AER*, Stahl and Wilson 1994 *JEBO*, 1995 *GEB*, Costa-Gomes, Crawford, and Broseta 2001 *Econometrica*, Camerer, Ho, and Chong 2004 *QJE*, Costa-Gomes and Crawford 2006 *AER*) suggests that subjects tend instead to follow rules of thumb that anchor beliefs in a simple model of others' instinctive reactions to the game and then adjust their beliefs via a small number of iterated best responses.

These rules of thumb—called "types"; no relation to private-information variables—are cognitively simple, and have strong intuitive appeal.

Subjects' thinking is typically heterogeneous, but their types are drawn from a stable population distribution concentrated on one to three best-response iterations.

The finite iteration of best responses by which people adjust their beliefs is common to all settings, although the number of iterations may vary.

The model of others' instinctive reactions people use to anchor their beliefs may take different forms in different settings:

● Uniform randomness (reflecting the principle of insufficient reason or payoff sampling uninformed by structure) in most normal-form games

● Attraction to salient labels or payoffs when these are important

● Truthfulness in games where players can communicate

The experimental evidence identifies a class of "level-*k*" or "cognitive hierarchy" ("CH") models that share the generality and much of the tractability of equilibrium analysis, but can often out-predict equilibrium.

Level-*k* types are rational in the sense of best-responding to some beliefs; they depart from equilibrium only in that their beliefs are derived from simplified, non-equilibrium models of other players.

Although level-*k*/CH models are alternatives to equilibrium analysis, they generalize equilibrium rather than replacing it.

Level-*k* type *k* (though not its CH counterpart beyond *k* = 1) respects *k*-rationalizability, the condition that corresponds in two-person games to the result of *k* rounds or iterated deletion of dominated strategies.

In sufficiently simple games, the low-level types that describe most subjects' behavior mimic equilibrium strategy choices, even though they deviate from equilibrium thinking.

But in more complex games, some or all such types may deviate systematically from equilibrium choices.

Importantly, a level-*k*/CH model not only predicts that such deviations will sometimes occur.

Given estimates of the population type frequencies, it also identifies which settings are likely to evoke deviations; what forms they will take; and with what frequencies.

Although level-*k*/CH models predict a sizeable fraction of deviations from equilibrium in many settings, they by no means predict all deviations in all interesting settings.

They seem to predict half or more of the deviations in a majority of normal-form settings.

This should not be disappointing: It is encouraging that such simple and tractable models can predict half or more of anything as elusive as deviations from equilibrium.

Moreover, the experimental results also suggest that the strategic thinking-related deviations that level-*k*/CH models do *not* predict have little discernable structure.

Thus, level-*k*/CH models generalize equilibrium analysis in a way that is likely to be useful in settings where deviations from equilibrium are important, while ignoring little that cannot be modeled as errors.

Experimental Evidence from Guessing and Other Normal-Form Games

“...professional investment may be likened 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.”

—John Maynard Keynes, *The General Theory of Employment, Interest, and Money*

“…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.”

—Warren Buffett’s intellectual hero Benjamin Graham (of Graham and Dodd’s *Security Analysis*), in Graham’s *The Intelligent Investor*

The Keynes and Graham quotations evoke simultaneous-move *n*-person guessing or “outguessing” games, possibly with multiple equilibria.

They concern games played without clear precedents.

The key issue they raise is anticipating others’ strategic responses: for Keynes to a “landscape” of personal judgments about prettiness, which is otherwise payoff-irrelevant; and for Graham to the psychology of a representative uninformed investor’s reaction to news.

Equilibrium is not very helpful in anticipating others’ responses in such settings.

Instead the quotations explicitly suggest thought processes in which players anchor beliefs in a model of others’ instinctive reactions and then iterate best responses a finite number of times, processes whose heterogeneity and finiteness closely resemble a level-*k*/CH model.

(Keynes’ “fourth, fifth and higher degrees” is more than the evidence suggests is realistic, but it may be only a coy reference to himself.)

Here I first discuss Nagel's (1995 *AER*); Ho, Camerer, and Weigelt's (1998 *AER*; "HCW"); and Bosch-Domènech et al.'s (2002 *AER*) analyses of *n*-person guessing games directly inspired by Keynes' beauty contest analogy, which give a simple introduction.

I then discuss Costa-Gomes and Crawford's (2006 *AER*; "CGC") analysis of two-person guessing games, whose design comes close to letting the data reveal subjects' thinking directly, without an econometric "middleman".

CGC's evidence and analysis are more precise than previous studies, but their conclusions are representative of the conclusions of most other studies of initial responses to normal-form games with neutral framing.

Other important experimental analyses of strategic thinking via eliciting initial responses to normal-form complete-information games include Stahl and Wilson (1994 *JEBO*, 1995 *GEB*); Costa-Gomes, Crawford, and Broseta (2001 *Econometrica*); Camerer, Ho, and Chong (2004 *QJE*); Chong, Camerer, and Ho (2005); and Costa-Gomes and Weizsäcker (2008 *REStud*).

Nagel's (1995); Ho, Camerer, and Weigelt's (1998); and Bosch-Domènech et al.'s (2002) experiments

In Nagel's and HCW's *n*-person guessing games, *n* subjects (*n* = 15-18 in Nagel, *n* = 3 or 7 in HCW) made simultaneous guesses between lower and upper limits (0 to 100 in Nagel, 0 to 100 or 100 to 200 in HCW).

In Bosch-Domènech et al. (2002 *AER*) essentially the same games were played in the field, by more than 7500 volunteers recruited from subscribers of the newspapers *Financial Times*, *Spektrum der Wissenchaft*, or *Expansión*.

In each case the subject who guessed closest to a target (*p* = 1/2, 2/3, or 4/3 in Nagel; *p* = 0.7, 0.9, 1.1, or 1.3 in HCW; and *p* = 2/3 in Bosch-Domènech et al.) times the group average guess won a prize.

There were several treatments, each with identical targets and limits for all players and games. The structures were publicly announced, to justify comparing the results with predictions based on complete information.

For definiteness, consider Nagel's leading treatment:

- 15-18 subjects simultaneously guessed between [0,100]

- The subject whose guess was closest to a target *p* (= 1/2 or 2/3, say), times the group average guess wins a prize, say $50

- The structure was publicly announced

If you are one of the few people in the world who have not already done so, please take a moment to decide what you would guess, in a group of non-game-theorists:

- if $p = 1/2$

- if $p = 2/3$

Nagel’s games have a unique equilibrium, in which all players guess 0.

The games are dominance-solvable, so the equilibrium can be found by iteratively eliminating dominated guesses.

For example, if $p = 1/2$:

- It’s dominated to guess more than 50 (because 1/2 × 100 ≤ 50).
- Unless you think that other people will make dominated guesses, it’s also dominated to guess more than 25 (because 1/2 × 50 ≤ 25).
- And so on, down to 12.5, 6.25, 3.125, and eventually to 0.

The rationality-based argument for this “all–0” equilibrium is stronger than many equilibrium arguments: it depends only on iterated knowledge of rationality, not on the assumption that players have the same beliefs.

However, even people who are rational are seldom certain that others are rational, or at least that others believe that others are rational.

Thus, they won’t (and shouldn’t) guess 0. But what do (should) they do?

Nagel’s and HCW’s subjects each played a game repeatedly, but their first-round guesses can be viewed as initial responses to a game played as if in isolation if they treated their own influences on future guesses as negligible, which is plausible for all but HCW’s 3-subject groups.

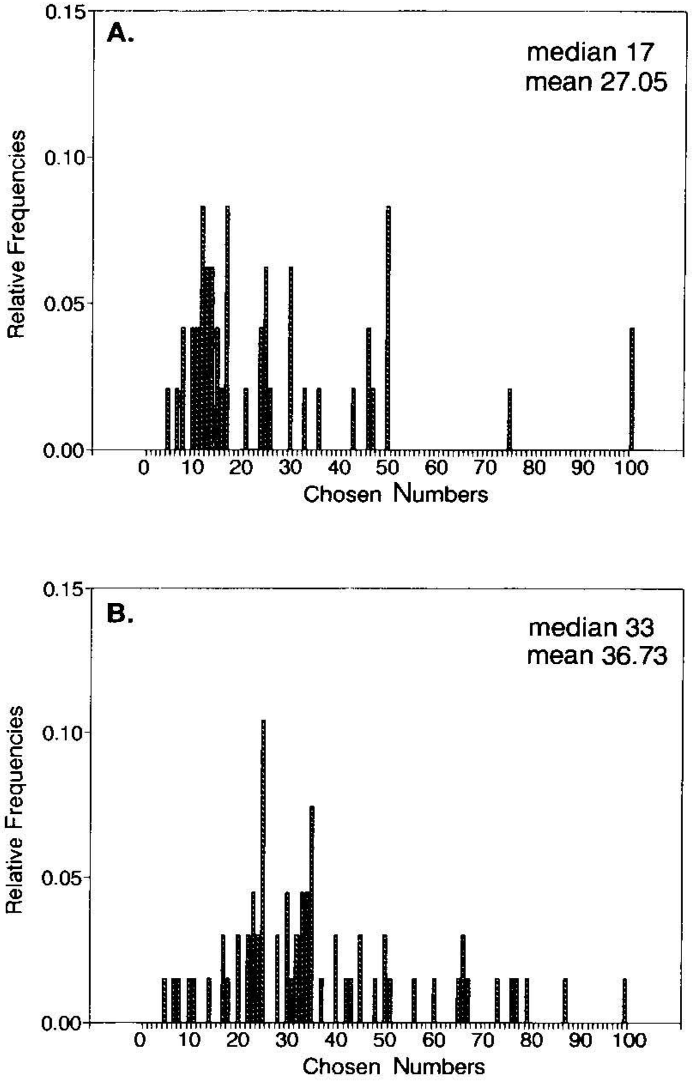
Bosch-Domènech et al.’s subjects played only once.

The results vividly illustrate the failure of equilibrium as a descriptive model of initial responses, and the heterogeneity and discreteness of strategic thinking.

Nagel’s subjects never made equilibrium guesses initially; HCW’s rarely did so, and Bosch-Domènech et al.’s (who had much more time to reflect, and who could consult with others) fairly rarely did so.

In each case most subjects’ initial guesses respected from 0 to 3 rounds of iterated dominance, in games where 3 to an infinite number are needed to reach equilibrium.

Here I reproduce part of Nagel’s Figure 1 (top *p* = 1/2, bottom *p* = 2/3) and Bosch-Domènech et al.’s Figure 1, which illustrate these points.



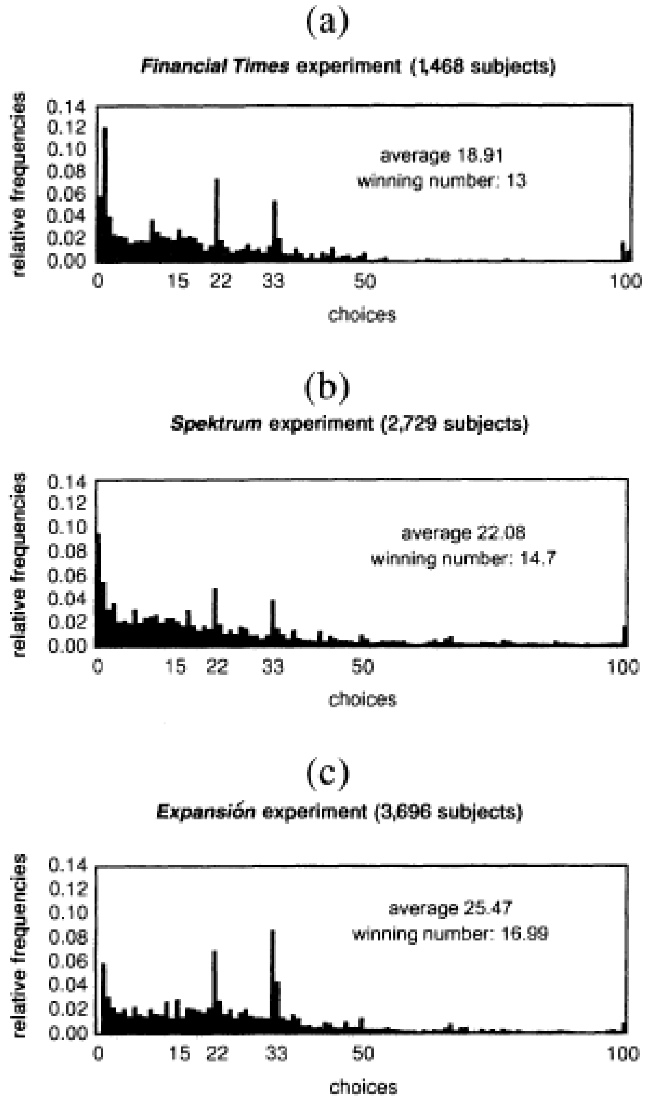


FIGURE 1. RELATIVE FREQUENCIES OF CHOICES
IN THREE NEWSPAPER EXPERIMENTS

These data resemble neither equilibrium plus noise nor “equilibrium taking noise into account” as in QRE (for any reasonable error distribution, though by Haile et al’s (2008 *AER*) result we could make the data an exact QRE for an unreasonable distribution).

The data do suggest that subjects’ deviations from equilibrium have a coherent structure.

The guess distributions have spikes that track $50p^k$ for k = 1, 2, 3 across treatments with various *p*s, respecting 0-3 rounds of iterated dominance.

Like the spectrograph peaks that foreshadow the existence of chemical elements, these spikes are evidence of a partly deterministic structure, one that is discrete and individually heterogeneous.

It is clear that no model that imposes homogeneity of strategic thinking (as most leading models but level-*k*/CH do) will do justice to behavior.

(Allowing heterogeneity is essential for the some explanations, including those proposed below for Kahneman’s Entry Magic and Huarongdao.)

Also, subjects do not respect indefinitely iterated dominance or indefinitely iterated best responses; instead their decisions respect *k*-rationalizability for at most small values of *k*.

But what about the spikes, whose consistency is the most remarkable part of the results?

Many theorists instinctively interpret Nagel's results as evidence that subjects explicitly performed finitely iterated dominance, the way we teach students to solve such games.

In this interpretation, which I will call *Dk*, a player does *k* rounds of iterated dominance for some small number, $k = 1$ or 2, and then best responds to a uniform prior over other players' remaining strategies (completing *k*-rationalizability by adding a specific selection): thus in Nagel's games *Dk* guesses $([0+100p^k]/2)p \equiv 50p^k$.

But there is another interpretation of the spikes, which has the same implications for choice behavior in Nagel’s games but which can differ in important ways in other settings, and which I shall argue is very likely the correct interpretation.

In this “level-*k*” interpretation, a player starts with a uniform prior *L0* over the strategy space and then iterates best responses *k* times, with $k = 1$, 2, or 3: thus in Nagel’s games *Lk+1* guesses $[(0+100)/2]p^{k+1}$, which equals $([0+100p^k]/2)p \equiv 50p^k$.

Note that it is *Lk*+1 that is *Dk*’s cousin, not *Lk*. The difference in indices is only a quirk of notation, without further significance.

Both *Lk*+1 and *Dk* yield *k*-rationalizable strategies, though not always the same ones in other games. In games without dominance *Dk*, $k = 1,2,\ldots$ coincides with *L1*.

Aside on specifying *L0*

These lectures focus mainly on two-person games.

But in *n*-person games like Nagel’s it matters whether *L0* is independent across players or correlated.

The limited evidence that is available (HCW and Costa-Gomes, Crawford, and Iriberri 2009 *JEEA*) suggests that most people have highly correlated (“representative agent”-like) models of others.

Accordingly, and in keeping with the literature, in analyzing Nagel’s data I take *L0* to directly model the distribution of all others’ average guess.

End of aside

The complete lack of separation of *Dk*’s and *Lk*+1’s guesses in Nagel’s design shows that the inference that subjects performed finitely iterated dominance is premature.

In other experiments, including some of HCW’s and Costa-Gomes, Crawford, and Broseta’s, *Dk*’s and *Lk*+1’s guesses are weakly separated, and the results are inconclusive on this point.

But in Costa-Gomes and Crawford’s experiment discussed next, *Dk*’s and *Lk*+1’s guesses are strongly separated, and the results very strongly favor *Lk+1* over *Dk* rules.

Thus, subjects’ guesses respect *k*-rationalizability for small *k* not because they explicitly perform iterated dominance, but because they follow rules that implicitly respect it.

Costa-Gomes and Crawford's (2006 *AER*) experiments

Nagel's and HCW's designs are distinguished by very large strategy spaces, which greatly increase the informativeness of their results.

But from the point of view of studying strategic thinking it is a weakness that each subject played only one game (although there was between-subjects variation across treatments).

Even though most subjects played their game repeatedly, their later choices confound strategic thinking with learning, so there was in effect only one observation per subject.

(Recall that first-round choices can still be viewed as initial responses to a game played as if in isolation if subjects treat their own influences on future choices as negligible.)

Even with very large strategy spaces, one observation yields limited information, and the results leave considerable ambiguity of interpretation regarding subjects' types.

By contrast, in Stahl and Wilson’s (1994 *JEBO*, 1995 *GEB*) and Costa-Gomes, Crawford, and Broseta’s (2001 *Econometrica*) designs, each subject played a series of different but related games, run in a way that suppresses learning and repeated-game effects.

However, in these experiments the games had small strategy spaces, with only two to four choices per player.

CGC’s design combines the variation through a series of games of those designs with the large strategy spaces of Nagel’s and HCW’s designs, greatly increasing the power of subjects’ choices to reveal their thinking.

Another advantage is that CGC’s design involves two-person games, in which a subject must predict the decisions of a partner who does not view the subject himself as a negligible part of the interaction; this fully engages strategic thinking in a way that games like Nagel’s do not.

Two-person games also avoid the “representative agent” ambiguity of interpretation noted in the aside above.

In CGC’s games subjects were randomly and anonymously paired to play a series of 16 different two-person guessing games, with no feedback.

The profile of a subject’s guesses in the 16 games forms a “fingerprint” that helps to identify his strategic thinking very precisely.

The design suppresses learning and repeated-game effects to elicit subjects’ initial responses, game by game, studying strategic thinking “uncontaminated” by learning.

(“Eureka!” learning was possible, but it was tested for and found to be rare.)

In CGC’s games each player has his own lower and upper limit, both strictly positive to make the games finitely dominance-solvable.

(Players are not required to guess between their limits. Guesses outside the limits are automatically adjusted up to the lower or down to the upper limit as necessary: a trick to enhance separation of information search implications, not important for this discussion.)

Each player also has his own target, and his payoff increases with the closeness of his guess to his target times the other’s guess.

The targets and limits vary independently across players and games, with targets both less than one, both greater than one, or “mixed”.

(In Nagel’s and HCW’s previous experiments the targets and limits were always the same for both players, and varied at most between subjects across treatments.)

CGC's guessing games have essentially unique equilibria ("essentially" due solely to the automatic adjustment), determined by players' lower (upper) limits when the product of targets is less (greater) than one.

Consider a game in which players' targets are 0.7 and 1.5, the first player's limits are [300, 500], and the second's are [100, 900].

The product of targets is 1.05 > 1, and the equilibrium is therefore determined by players' upper limits.

In equilibrium the first player guesses his upper limit 500, but the second player guesses 750 (= 500 × his target 1.5), below his upper limit 900.

No guess is dominated for the first player, but any guess outside [450, 750] is dominated for the second player. Given this, any guess outside [315, 500] is iteratively dominated for the first player. Given this, any guess outside [472.5, 750] is dominated for the second player, and so on until the equilibrium at (500, 750) is reached after 22 iterations.

The discontinuity of the equilibrium correspondence when the product of targets is one enhances the separation of equilibrium from other types.

The design stress-tests equilibrium in that it includes games that differ mainly in whether the product of targets is slightly greater or slightly less than one, a difference equilibrium responds to much more strongly than behaviorally plausible nonequilibrium types do.

It also yields other interesting results, discussed in the paper.

CGC'S data analysis

As suggested by previous work, CGC's data analysis assumed that each subject's guesses were determined, up to logit errors, by a single type in all 16 games. This assumption was tested and found reasonable.

Most of the analysis restricted attention to a list of plausible types:

● *L0*, *L1*, *L2*, and *L3* as defined above, with *L0* uniform random

● *D1* and *D2* as defined above

● *Equilibrium*, which makes its equilibrium decisions

● *Sophisticated*, which best responds to the probability distributions of others' decisions, estimated from the observed frequencies (an ideal, included to learn if any subjects have an understanding of others' decisions that transcends mechanical rules.)

The restriction to this list was also tested and found to be a reasonable approximation to the support of subjects' decision rules.

CGC’s large strategy spaces and the independent variation of targets and limits across games greatly enhance the separation of types’ implications, to the point where many subjects’ types can be precisely identified from their guessing “fingerprints”:

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

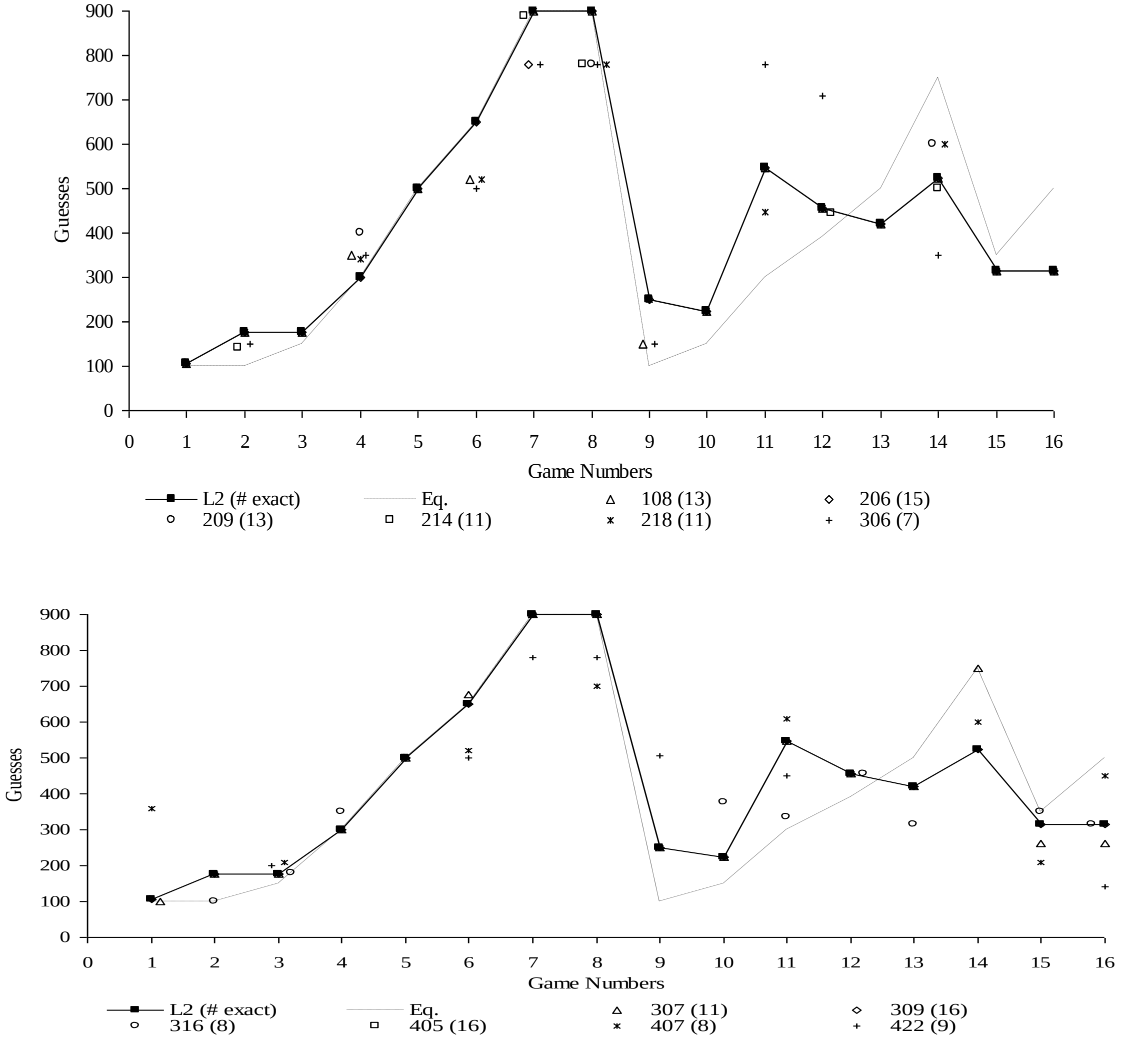
	L1	***L2***	***L3***	***D1***	***D2***	***Eq.***	***Soph.***
1	**600**	**525**	**630**	**600**	**611.25**	**750**	**630**
2	**520**	**650**	**650**	**617.5**	**650**	**650**	**650**
3	**780**	**900**	**900**	**838.5**	**900**	**900**	**900**
4	**350**	**546**	**318.5**	**451.5**	**423.15**	**300**	**420**
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**

Of the 88 subjects in CGC's main treatments, 43 made guesses that complied *exactly* (within 0.5) with one type's guesses in from 7 to 16 of the games:

20 *L1*, 12 *L2*, 3 *L3*, and 8 *Equilibrium*.

For example, CGC's Figure 2 (next slide) shows the "fingerprints" of the 12 subjects whose guesses conformed most closely to *L2*'s.

Of these subjects' 192 guesses, 138 (72%) were exact *L2* guesses.



CGC's Figure 2. "Fingerprints" of 12 Apparent *L2* Subjects (Only deviations from *L2*'s guesses are shown.)

The size of CGC’s strategy spaces, with 200 to 800 possible exact guesses in each of 16 different games, makes exact compliance powerful evidence for a type:

If a subject chooses 525, 650, 900 in games 1-3, intuitively and econometrically we already “know” he’s an *L2*.

(By contrast, there are usually many possible reasons for choosing one of the strategies in a small matrix game; and even in Nagel’s games, rules as cognitively disparate as *Dk* and *Lk*+1 yield identical decisions.)

Further, because CGC’s definition of *L2* builds in risk-neutral, self-interested rationality, we also know that a subject’s deviations from equilibrium are caused not by irrationality, risk aversion, altruism, spite, or confusion, but by his simplified model of others.

(Even so, doubts remain about the subjects with high exact compliance with *Equilibrium*, who seem to follow types that only mimic equilibrium in games with targets both less than one or greater than one.)

That the level-*k* model is *directly* suggested by the data for half of CGC's subjects (rather than via data-fitting exercises) is an important advantage over alternatives.

CGC's other 45 subjects made guesses that conformed less closely to a type.

But for all but 14 of them, violations of simple dominance were comparatively rare (less than 20%, versus 38% for random guesses), suggesting that their behavior was coherent, even though less well described by the types.

And econometric estimates of their types are concentrated on *L1*, *L2*, *L3*, and *Equilibrium* in roughly the same proportions.

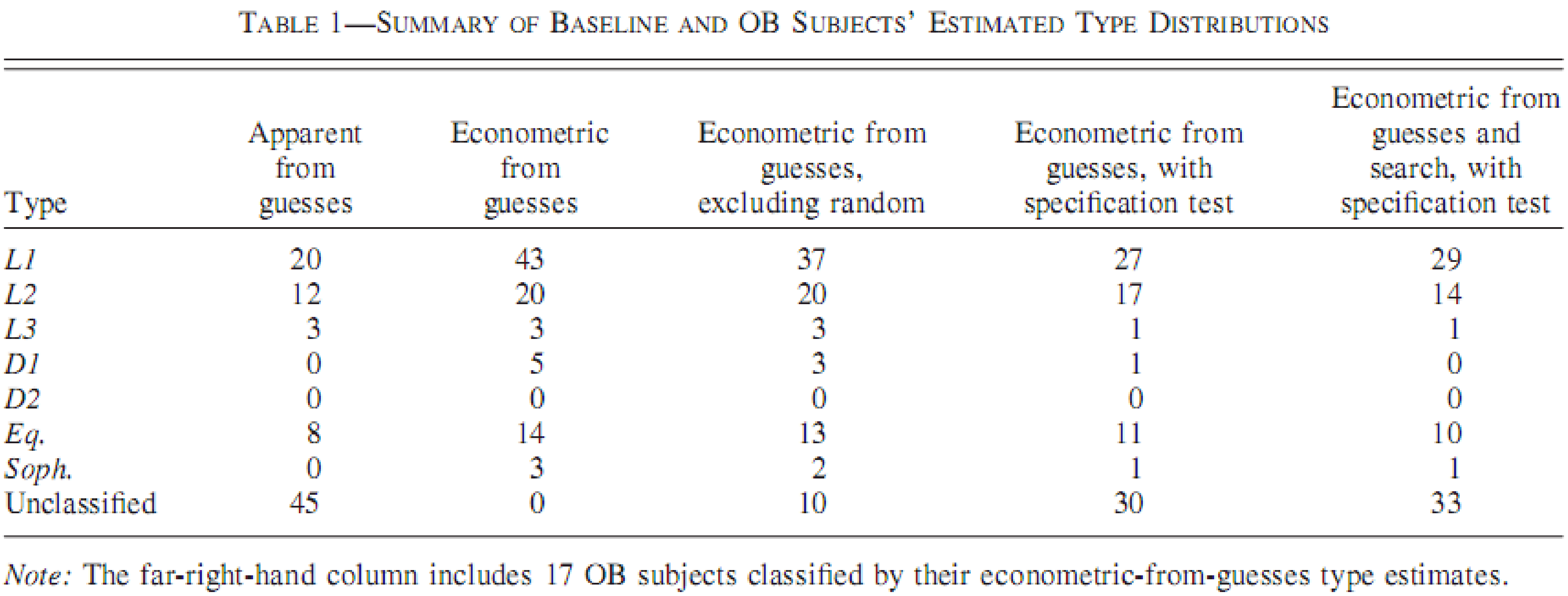


TABLE 1—SUMMARY OF BASELINE AND OB SUBJECTS' ESTIMATED TYPE DISTRIBUTIONS

Type	Apparent from guesses	Econometric from guesses	Econometric from guesses, excluding random	Econometric from guesses, with specification test	Econometric from guesses and search, with specification test
L1	20	43	37	27	29
L2	12	20	20	17	14
L3	3	3	3	1	1
D1	0	5	3	1	0
D2	0	0	0	0	0
Eq.	8	14	13	11	10
Soph.	0	3	2	1	1
Unclassified	45	0	10	30	33

Note: The far-right-hand column includes 17 OB subjects classified by their econometric-from-guesses type estimates.

CGC's Figure 1.

Lessons for modeling strategic behavior in initial responses

CGC’s analysis reinforces and sharpens the conclusions of previous analyses:

- No model that imposes homogeneity of strategic thinking will do justice to subjects’ behavior

- Subjects do not respect indefinitely iterated dominance or best responses; their decisions respect *k*-rationalizability only for small *k*

- There are few if any *Dk* subjects; people respect iterated dominance to the extent that their *Lk* types do, not because they explicitly perform it. (This is strongly reinforced by CGC’s data on subjects’ searches for hidden payoff information and their data on “robot/trained subjects”.)

- There are few if any *Sophisticated* subjects. (The jury is still out on the extent to which this conclusion generalizes.)

CGC’s results show that a hybrid level-*k* or CH model with a uniform random *L0* and only *L1*, *L2*, *L3*, and possibly *Equilibrium* subjects explains a large fraction of subjects’ deviations from equilibrium.

CGC’s results could still be domain-specific, but they are consistent with most other results from different settings, just more precise.

Further, although about half of CGC’s subjects’ deviations from equilibrium remain unexplained by a hybrid level-*k* or CH model, the specification test suggests that the deviations have little discernable structure.

Thus it may still be optimal for a modeler to treat the remaining deviations as errors, and the part of the structure that can be identified may provide a basis for unbiased modeling of initial responses to games.

Application: M. M. Kaye’s Far Pavilions: Responding to Payoff Asymmetries in Outguessing Games

“…ride hard for the north, since they will be sure you will go southward where the climate is kinder….”

> —Koda Dad (played by Omar Sharif in the HBO miniseries) to Ash/Ashok in M.M. Kaye’s *The Far Pavilions* (1978, p. 97)

I now consider an example that illustrates applications of level-*k* models.

Early in *The Far Pavilions*, the main male character, Ash/Ashok, is trying to escape from his pursuers along a north-south road.

Both Ash and his pursuers must choose between north and south. Although Ash moves first, the pursuers must make their choice irrevocably before they learn Ash’s choice, so their choices are strategically simultaneous.

South is warm, but north are the Himalayas, with winter coming.

Imagine that if the pursuers catch Ash, they gain two units of payoff and Ash loses two, and that both the pursuers and Ash gain one extra unit for choosing South, whether or not Ash is caught.

This yields the payoff matrix:

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

***Far Pavilions* Escape**

Examples like this are as common in experimental game theory as they are in fiction, but fiction sometimes more clearly reveals the thinking behind a decision.

Ash's mentor Koda Dad advises Ash to ride north, for the reason given in the quotation:

> "…ride hard for the north, since they will be sure you will go southward where the climate is kinder…."

Ash overcomes his fear of freezing and follows this advice, the pursuers unimaginatively go south, and Ash escapes.

Koda Dad is advising Ash to choose the *L3* response to a uniform random *L0*.

(*L3* ties my personal best *k* for a clearly explained level-*k* type in fiction. I suspect even postmodern fiction may have nothing clearly higher than *L3*: it wouldn’t be credible.)

If the pursuers expect Ash to go south because it’s “kinder”, they must be modeling Ash as an *L1* response to a uniform random *L0*.

For, the payoff asymmetry on which this inference rests is decisive only if north and south do not differ in the probability of being caught, which is more important.

Thus, Koda Dad must be modeling the pursuers as *L2* and advising Ash to choose the *L3* response to a uniform random *L0*.

We could take the inference that Ash will go south because it’s “kinder” literally as a best response to a uniform random *L0*.

But there is a behaviorally more plausible interpretation in which the inference is Ash’s model of the pursuers’ model of Ash’s instinctive reaction ignoring strategic considerations, and given this, plausibly based on the principle of insufficient reason.

In a more complex game, a uniform random *L0* could plausibly approximate random sampling of payoffs unstratified by the other player’s strategy choices.

How does the level-*k* model compare in predictive success with an equilibrium model?

Escape has a unique equilibrium in mixed strategies, in which

$$3p + 1(1 - p) = 0p + 2(1 - p) \text{ or } p = 1/4, \text{ and}$$
$$-1q + 1(1 - q) = 0q - 2(1 - q) \text{ or } q = 3/4.$$

Thus Ash's Pr{South}, $p^* = 1/4$, and the Pursuers' Pr{South}, $q^* = 3/4$.

This equilibrium responds to the payoff asymmetry between South and North in a decision-theoretically intuitive way for Pursuers (because $q = 3/4 >$ the 1/2 of equilibrium without the payoff asymmetry), but counterintuitively for Ash (because $p = 1/4 < 1/2$).

In equilibrium the novel's observed outcome {Ash North, Pursuers South} has probability $(1 - p^*)q^* = 9/16$: much better than a random 25%.

By contrast, the level-*k* model implies decisions as follows:

Type	**Ash**	**Pursuers**
L0	uniformly random	Uniformly random
L1	South	South
L2	North	South
L3	North	North
L4	South	North
L5	South	South

***Lk* types' decisions in *Far Pavilions* Escape**

Thus the level-*k* model predicts the outcome provided that Ash is either *L2* or (as we know from the quotation) *L3*, and the Pursuers are either *L1* or (as Koda Dad believes) *L2*.

Of course, applications seldom come with an omniscient narrator telling us how players think.

Even so, if the game is clearly defined and we have data, we can specify a level-*k* model, derive its implications, and use them to estimate the type frequency distribution.

Alternatively, we can calibrate the model using previous estimates from similar settings.

Suppose, for example, that we calibrate a level-*k* model by assuming that each player role in Escape is filled from a plausible 50-30-20 mixture of *L1*s, *L2*s, and *L3*s, and that there are no errors.

Then the predicted frequency with which Ash goes North is 1/2 and the frequency with which the Pursuers go South is 4/5.

Assuming independence, this implies that the observed outcome {Ash North, Pursuers South} has probability 2/5: less than the equilibrium predicted frequency of 9/16, but still better than a random 25%.

More importantly, the level-*k* model gracefully explains a puzzling divergence between observed aggregate behavior patterns and equilibrium predictions.

In games like Escape and closely related perturbed Matching Pennies games, the unique mixed-strategy equilibrium responds to the payoff asymmetry between South and North in a decision-theoretically intuitive way for the pursuers' role ($q^* = 3/4 > 1/2$, the probability with which pursuers go south in the analogous game with no north-south payoff asymmetry); but in a counterintuitive way for Ash's role ($p^* = 1/4 < 1/2$).

Yet experimental subjects' aggregate choices in initial responses to games like this reflect decision-theoretic intuition in both player roles. (Ash's counterintuitive choice would not contradict this pattern if he were a subject, because his revealed type is in the minority.)

In such games the level-*k* and CH models' predictions "quasi-purify" something roughly like a mixed-strategy equilibrium via the predictable heterogeneity of players' strategic thinking, while avoiding some implausible implications of equilibrium.

Application: Groucho’s Curse: Zero-Sum Betting and Auctions with Incomplete Information

“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 (1959, p. 321), Telegram to the Beverly Hills Friar’s Club

“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)

Although most laboratory evidence on strategic thinking comes from symmetric-information designs, most field evidence and applications involve settings with informational asymmetries.

It is therefore important to extend what can be learned about strategic thinking in complete-information games to incomplete-information games.

I now discuss laboratory and field evidence on games with informational asymmetries, focusing on cases where the games allow clear inferences about strategic thinking.

Experiments on zero-sum betting

Experiments on zero-sum betting build on Milgrom and Stokey’s (1982 *JET*) no-trade theorem, which shows that if traders are weakly risk-averse and have consistent beliefs, and the initial allocation is Pareto-efficient relative to the information available, then even if traders receive new private information, no weakly mutually beneficial trade is possible; and if traders are strictly risk-averse, no trade at all is possible.

For, any such trade would make it common knowledge that both traders had benefited, contradicting the hypothesis that the original allocation was Pareto-efficient.

This result has been called the Groucho Marx theorem because its logic resembles that of our Marx quotation.

By contrast with the conclusions of the theorem, speculative zero-sum trades are common in real markets. This fact has a number of possible explanations, of which one is nonequilibrium strategic thinking.

The experiments by Brocas, Carrillo, Camerer, and Wang (2009) (see also Camerer, Ho, and Chong 2004 *QJE*, Section VI, and Rogers, Palfrey, and Camerer 2009 *JET*) have the control required to distinguish between such explanations and those based on factors like hedging or joy of gambling.

Brocas et al.'s design used simple three-state betting games, including this one:

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

Zero-Sum Betting Game

Each of two players, 1 and 2, is given information about which of three ex ante equally likely states has occurred, A, B, or C. As indicated in Figure 1, player 1 learns either that the state is {A or B} or that it is C; player 2 learns that the state is A or that it is {B or C}.

The rules of the game and the information structure were publicly announced, with the goal of inducing common knowledge.

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

Zero-Sum Betting Game

Once informed, the players choose simultaneously between two decisions: Bet or Pass.

A player who chooses Pass earns 10 no matter what the state. If one chooses Bet while the other chooses Pass, they both earn 10.

If both players choose Bet, they get one of the payoffs in the table, depending on which state has occurred.

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

Zero-Sum Betting Game

This game has a unique trembling-hand perfect Bayesian equilibrium.

In this equilibrium, player 1 told C will Bet because 20 > 10, and player 2 told A will Pass because 0 < 10.

Given this, player 1 told {A or B} will Pass, because player 2 will Pass if told A, so betting given {A or B} yields player 1 at most 5 < 10.

Given this, player 2 will Pass if told {B or C}, because player 1 will Pass if told {A or B}, so betting given {B or C} yields player 2 at most 5 < 10.

This covers all contingencies and completes the characterization of equilibrium, which shows that the game is weakly dominance-solvable in three rounds. No betting takes place in equilibrium in any state.

Despite this clear equilibrium conclusion, in Brocas et al.’s and several similar experiments approximately half of the subjects bet.

To explain this Brocas et al. proposed a level-*k* model with a specification like those discussed above, but with *L0* adapted to allow for incomplete information.

Following Camerer, Ho, and Chong (2004 iQJE, Section VI) and Crawford and Iriberri (2007 *Econometrica*), Brocas et al. assumed that *L0* bids uniformly randomly, independent of its private information.

In judging this specification, bear in mind that *L0* is meant to describe a player’s model of the instinctive starting point of others’ strategic thinking, from which point of view it is behaviorally plausible that *L0* ignores others’ private information, which it does not observe.

As in previous level-*k* analyses, Brocas et al. took their *L1* to best respond to their *L0,* and their *L2* to best respond to their *L1*.

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

Zero-Sum Betting Game

Following Crawford and Iriberri (2007 *Econometrica*), call an *L1* that best responds to a random *L0* “random *L1*” even though it is not itself random; and call an *L2* that best responds to a random *L1* “random *L2*”.

Given this, random *L1* player 1s Bet if told {C} because it yields 20 > 10 if player 2 Bets, a random *L0* player 2 bets with probability one-half in either contingency, and Betting is otherwise costless.

Unlike in equilibrium, random *L1* player 1s Bet if told {A or B} because it yields 25 in state {A} and a random *L0* player 2 bets with probability one-half in {A}, it yields 5 in state {B} and a random *L0* player 2 Bets with probability one-half in {B or C}, the two states are equally likely ex ante, so Betting if told {A or B} yields expected payoff (25 + 5)/2 = 15 > 10.

Random *L1* player 2s Pass if told {A}, because it yields 0 < 10.

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

Zero-Sum Betting Game

Unlike in equilibrium, random *L1* player 2s Bet if told {B or C}, because it yields 30 in state {B} and a random *L0* player 1 Bets with probability one-half in {A or B}, it yields 5 in state {C} and a random *L0* player 1 Bets with probability one-half in {C}, the two states are equally likely ex ante, so Betting if told {B or C} yields expected payoff (30 + 5)/2 = 17.5 > 10.

Thus, if all subjects were random *L1s*, 100% of player 1s and 67% of player 2s would Bet, too much in each role to fit the aggregate data; and betting would occur only in states B and C, also not true in the data.

Brocas et al.’s data analysis finds clusters corresponding to random *L1*s, *L2*s, and *L3*s (who correspond here to *Equilibrium* players), and a cluster of apparently irrational players, which mixture of types fits much better.

Auction experiments

There is a rich literature on sealed-bid incomplete information auction experiments, which despite similar goals and methods has developed largely independently of the literature on game experiments.

In sealed-bid auction experiments, subjects' initial responses tend to exhibit overbidding, relative to the risk-neutral Bayesian equilibrium, in first- or second-price, independent-private-value or common-value auctions (Kagel and Levin 1986 *AER*, Goeree, Holt, and Palfrey 2002 *JET*).

The literature on auction experiments has proposed different explanations of overbidding in private- and common-value auctions: "joy of winning" and/or risk-aversion for private-value auctions, and the winner's curse for common-value auctions.

Moreover, these explanations are only loosely related to explanations that have been proposed for deviations from equilibrium in other games.

Kagel and Levin (1986 *AER*) and Eyster and Rabin (2005 *Econometrica*) sought to unify the explanations of nonequilibrium behavior in auctions and other games.

Kagel and Levin formalize the intuition behind the curse in models in which “naïve” bidders do not adjust their value estimates for the information revealed by winning (essentially, random *L1* bidding), but otherwise follow equilibrium logic.

Eyster and Rabin’s notion of “cursed equilibrium”, in which people underestimate the correlation between others’ decisions and private information but otherwise behave as in equilibrium, generalizes Kagel and Levin’s model to allow intermediate levels of value adjustment, ranging from standard equilibrium with full adjustment to “fully-cursed” equilibrium with no adjustment.

Both models allow players to deviate from equilibrium only to the extent that they do not draw correct inferences from the outcome. Thus their predictions coincide with equilibrium in games in which such inferences are not relevant, and they do not help to explain non-equilibrium behavior in independent-private-value auctions.

Crawford and Iriberri (2007 *Econometrica*) propose a level-*k* analysis of behavior in the classic auction experiments of Kagel and Levin 1986 *AER* and others, which provides an alternative way to unify the explanation of results for initial responses in auction experiments, without invoking joy of winning or risk aversion.

Crawford and Iriberri's analysis explores the robustness of equilibrium auction theory to failures of the equilibrium assumption and makes a connection between experiments on auctions and strategic thinking.

The key issue is how to specify *L0*; there are two possibilities here:

- *Random L0*, as in Brocas et al.'s analysis of zero-sum betting, bids uniformly on the interval between the lowest and highest possible values (even if above own value, given that *L0* represents a player's model of others' instinctive, nonstrategic responses to the game).

- *Truthful L0*, which is meaningful in auctions though not in all incomplete-information games, bids its expected value conditional on its own signal.

Crawford and Iriberri build separate type hierarchies on these *L0*s:

Random (*Truthful*) *Lk* is defined by iterating best responses from *Random* (*Truthful*) *L0*.

In their empirical analysis, they allow each subject to be one of the types, from either hierarchy.

For a given *Lk* type, as in an equilibrium analysis, the optimal bid must take into account value adjustment for the information revealed by winning (only in common-value auctions), and the bidding trade-off between the higher price paid if the bidder wins and the probability of winning (only in first-price auctions).

Crawford and Iriberri show that their level-*k* model allows a tractable characterization of these aspects of the bidder’s problem, which closely parallels the equilibrium characterization (but yields different results).

With regard to value adjustment, Random *L1* does not condition on winning because Random *L0* bidders bid randomly, independent of their values; thus Random *L1* is “fully cursed” in Eyster and Rabin’s sense.

The other types do condition on winning in various ways, but this conditioning tends to make bidders’ bids strategic substitutes, in that the higher others’ bids are, the greater the (downward) adjustment.

Thus, to the extent that Random *L1* overbids, Random *L2* tends to underbid (relative to equilibrium): If it’s bad news that you beat equilibrium bidders, it’s even worse news that you beat overbidders.

The bidding tradeoff, by contrast, can go either way, as it can in an equilibrium analysis.

The question, empirically, is whether an estimated mixture of Random *L1* overbidding and Random *L2* underbidding fits the data from auction experiments better than alternative models.

In three of four leading cases, a level-*k* model does better than equilibrium plus noise, cursed equilibrium, and/or logit QRE.

For the remaining case (Kagel and Levin's first-price auction), the most flexible cursed equilibrium specification has a small advantage.

Except in Kagel and Levin's second-price auctions, the estimated type frequencies are similar to those found in other experiments:

Crawford and Iriberri estimated low or zero frequencies of Random and Truthful *L0*, large frequencies (59-65%) of random *L1* bidders, and much smaller but significant frequencies of random *L2* (4-9%), truthful *L1* (9-18%), and truthful *L2* (1-16%).

Application: Kahneman’s Entry Magic: Coordination via Symmetry-Breaking

“…to a psychologist, it looks like magic.”

—Kahneman 1988, quoted in Camerer, Ho, and Chong (2004 *QJE*)

Kahneman’s “magic” refers to the fact that subjects in his own and others’ (Rapoport et al. 1998 and Rapoport and Seale 2002) market-entry experiments made choices surprisingly close to equilibrium in the aggregate.

(Thus Kahneman should have said “…to a *game theorist*, it looks like magic.”)

Another interesting feature was that subjects achieved systematically better coordination ex post than in the natural equilibrium benchmark.

In these experiments, *n* subjects choose simultaneously between entering (“In”) and staying out (“Out”) of a market with given capacity.

In yields a given positive profit if no more subjects enter than a given market capacity; but a given negative profit if too many enter.

For simplicity, Out yields 0 profit, no matter how many subjects enter.

In these games, efficient coordination requires breaking the symmetry of players’ roles.

But because players cannot distinguish their roles, it is not sensible to predict systematic differences across roles in behavior.

Thus, the natural equilibrium benchmark is the unique, symmetric mixed-strategy equilibrium, in which each player enters with a given probability that makes all players indifferent between In and Out.

This mixed-strategy equilibrium yields an expected number of entrants approximately equal to market capacity, but there is a positive probability that either too many or too few will enter.

Subjects came surprisingly close to this equilibrium benchmark.

But they also tended to have systematically better coordination ex post (number of entrants stochastically closer to market capacity) than in the equilibrium.

A level-k Analysis of Two-Person Entry/Battle of the Sexes Games

Camerer, Ho, and Chong (2004 *QJE*, Section III.C) explain Kahneman’s magic via a CH, in which the heterogeneity of strategic thinking allows some players to mentally simulate others’ entry decisions and accommodate them.

The more sophisticated players become somewhat like Stackelberg followers, which in entry games yields coordination benefits for all.

Here I illustrate their analysis in a two-person Battle of the Sexes game, which is like a two-person market-entry game with capacity one, and which makes the central points as simply as possible.

For simplicity, I also substitute a level-*k* model for their CH model.

The analysis illustrates the importance of the structured heterogeneity of strategic thinking a level-*k* model allows.

Consider a two-person Battle of the Sexes game with *a* > 1:

	In	**Out**
In	**0** **0**	**1** ***a***
Out	***a*** **1**	**0** **0**

Battle of the Sexes

The unique symmetric equilibrium is in mixed strategies, with $p \equiv \Pr\{\text{In}\} = a/(1+a)$ for both players.

The mixed-strategy equilibrium expected coordination rate is $2p(1 - p) = 2a/(1+a)^2$, and players' equilibrium expected payoffs are $a/(1+a)$.

This expected coordination rate is maximized when $a = 1$, where it takes the value ½.

With $a > 1$ these expected payoffs $a/(1+a) < 1$: worse for each player than his worst pure-strategy equilibrium. As $a \to \infty$, the expected coordination rate $2a/(1 + a)^2 \to 0$ like $1/a$.

Now consider a level-*k* model in which each player follows one of four types, *L1*, *L2*, *L3*, or *L4*, with each role filled by a draw from the same distribution.

For simplicity assume the frequency of *L0* is 0, and that *L0* chooses uniformly randomly, with Pr{In} = Pr{Out} = 1/2.

Type pairings	***L1***	***L2***	***L3***	***L4***
L1	**In, In**	**In, Out**	**In, In**	**In, Out**
L2	**Out, In**	**Out, Out**	**Out, In**	**Out, Out**
L3	**In, In**	**In, Out**	**In, In**	**In, Out**
L4	**Out, In**	**Out, Out**	**Out, In**	**Out, Out**
Outcomes in Battle of the Sexes				

*L1*s mentally simulate *L0*s' random decisions and best respond, thus, with *a* > 1, choosing In; *L2*s choose Out; *L3*s choose In; and *L4*s choose Out.

The predicted outcome distribution is determined by the outcomes of the possible type pairings and the type frequencies.

If both roles are filled from the same distribution, players have equal ex ante payoffs, proportional to the expected coordination rate.

L3 behaves like *L1*, and *L4* like *L2*.

Lumping *L1* and *L3*, and also *L2* and *L4*, together and letting v denote their total probability, the expected coordination rate is $2v(1 – v)$.

This is maximized at $v = ½$, where it takes the value ½.

Thus for v near ½, which is behaviorally plausible, the coordination rate is near ½.

For more extreme values the rate is worse, → 0 as $v \to 0$ or 1.

But because the equilibrium coordination rate of $2a/(1 + a)^2 \to 0$ like $1/a$, even for moderate values of a, the level-*k* coordination rate is higher.

The level-*k* analysis suggests a view of tacit coordination profoundly different from the traditional view, and illustrates the importance of the heterogeneity of strategic thinking the model allows.

Equilibrium and selection principles such as risk- or payoff-dominance (Harsanyi and Selten 1987) play no direct role in players' thinking.

Coordination, when it occurs, is an almost accidental (though statistically predictable) by-product of the use of non-equilibrium decision rules.

Even though players' decisions are simultaneous and there is no communication or observation of the other's decision, the heterogeneity of strategic thinking allows more sophisticated players such as *L2*s to mentally simulate the decisions of less sophisticated players such as *L1*s and accommodate them, just as Stackelberg followers would.

This mental simulation doesn't work perfectly, because an *L2* is as likely to be paired with another *L2* as an *L1*. Neither would it work if strategic thinking were homogeneous. But it's very surprising that it works at all.

Application: Yushchenko and Lake Wobegon: Non-neutral Framing in Outguessing Games

> "Any government wanting to kill an opponent…would not try it at a meeting with government officials."
>
> > —comment, quoted in Chivers (2004), on the poisoning of Ukrainian presidential candidate—now ex-president—Viktor Yushchenko.
>
> "…in Lake Wobegon, the correct answer is usually 'c'."
>
> > —Garrison Keillor (1997) on multiple-choice tests (quoted in Attali and Bar-Hillel 2003)

Both quotations refer to simultaneous-move zero-sum two-person games with unique mixed-strategy equilibria.

In the first, the players are a government assassin choosing one of several dinners at which to try to poison Yuschenko, only one of which is with officials of the government suspected of wanting to poison him; and an investigator who has the resources to check only one of the dinners.

In the second, the players are a test designer deciding where to hide the correct answer and a clueless test-taker trying to guess the hiding place.

Although there is nothing as uniquely salient as the dinner with government officials, psychologists think that with four possible answers, both the a and d end locations and location c are inherently salient (with the jury still out on which is more salient; see Christenfeld 1995 and Rubinstein, Tversky, and Heller 1996).

In each case the key issue is how players react to framing of decisions that is non-neutral but does not directly affect payoffs.

Equilibrium in zero-sum two-person games leaves no room for such framing to affect outcomes, but people often react to it anyway.

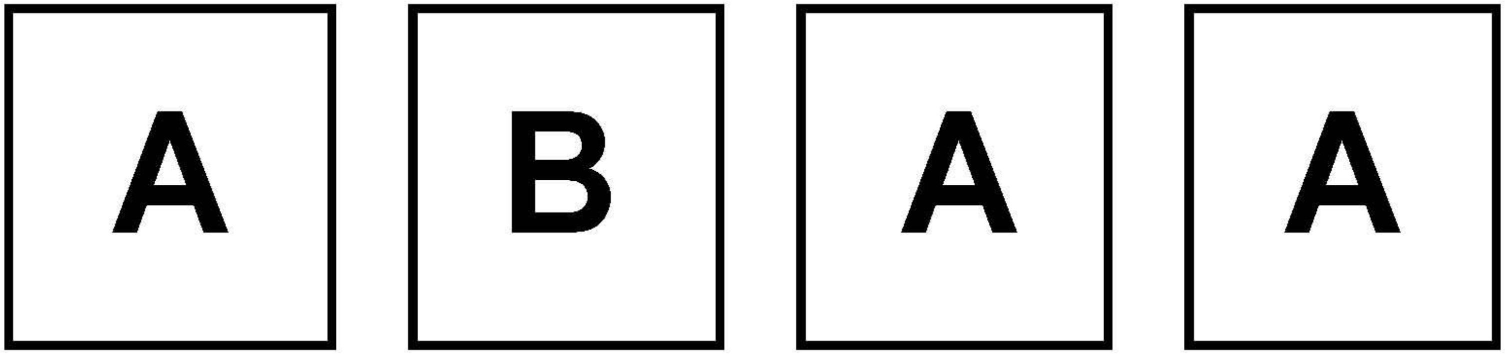
The thinking reflected by the quotations is plainly strategic, but non-equilibrium:

To the first, for example, any game theorist worth his salt would respond, “If that’s what people think, a meeting with government officials is exactly where *I* would try to poison Yushchenko.”

We will see that the quotation can be understood as a thought process in which a player anchors his beliefs in an instinctive reaction to the salience of the dinner with government officials and then iterates best responses a small number of times.

Consider Rubinstein, Tversky, and Heller’s 1993, 1996, 1998-99 (“RTH”) experiments with zero-sum, two-person “hide-and-seek” games with non-neutral framing of locations, analyzed by Crawford and Iriberri (2007 *AER*).

A typical seeker’s instructions (a hider’s instructions are analogous): *Your opponent has hidden a prize in one of four boxes arranged in a row. The boxes are marked as shown below: 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?*



RTH’s framing of the hide-and-seek game is non-neutral in two ways:

- The “*B*” location is distinguished by its label.
- The two “*end A*” locations may be inherently focal.

This gives the “*central A*” location its own brand of uniqueness as the “least salient” location.

Mathematically this “negative” uniqueness is analogous to the “positive” uniqueness of “*B*”.

However, its psychological effects will be seen to be completely different.

RTH’s design is important as a tractable abstract model of a non-neutral cultural or geographic frame, or “landscape.”

Hide-and-seek games are often played on such landscapes, although traditional game theory rules out any influence of the landscape by fiat.

This is well illustrated by the Yuschenko and Lake Wobegon quotations.

Yuschenko’s meeting with government officials is analogous to RTH’s B, although in that example there’s nothing like RTH’s end locations.

With four possible choices arrayed left to right in the zero-sum game between a test designer deciding where to hide the correct answer and a clueless test-taker trying to guess where it is, the Lake Wobegon example is very close to RTH’s design.

RTH’s hide-and-seek game has a clear equilibrium prediction, which leaves no room for framing to systematically influence the outcome.

The traditional theory of zero-sum two-person games is the strongpoint of noncooperative game theory, where the normative arguments for playing equilibrium strategies are immune most counterarguments.

Yet framing has a strong and systematic effect in RTH’s experiments, qualitatively the same around the world, with *Central A* (or its analogs in other treatments, as explained in the paper) most prevalent for hiders (37% in the aggregate) and even more prevalent for seekers (46%).

In this game one might argue that deviations do not violate the theory because any strategy is a best response to equilibrium beliefs.

But systematic deviations of aggregate frequencies from equilibrium probabilities must (with high probability) have a systematic cause.

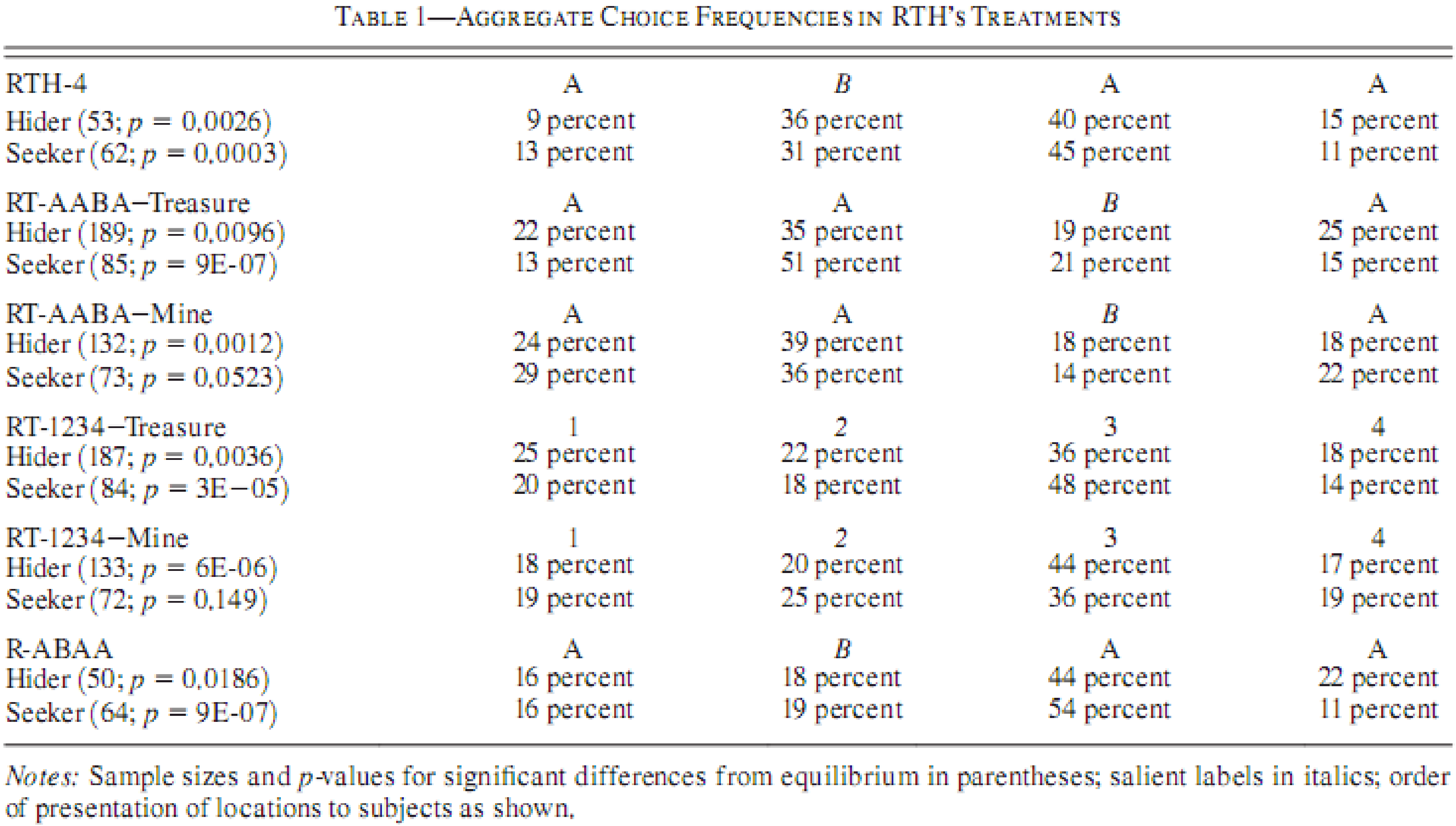


TABLE 1—AGGREGATE CHOICE FREQUENCIES IN RTH'S TREATMENTS

RTH-4	A	*B*	A	A
Hider (53; p = 0.0026)	9 percent	36 percent	40 percent	15 percent
Seeker (62; p = 0.0003)	13 percent	31 percent	45 percent	11 percent
RT-AABA–Treasure	A	A	*B*	A
Hider (189; p = 0.0096)	22 percent	35 percent	19 percent	25 percent
Seeker (85; p = 9E-07)	13 percent	51 percent	21 percent	15 percent
RT-AABA–Mine	A	A	*B*	A
Hider (132; p = 0.0012)	24 percent	39 percent	18 percent	18 percent
Seeker (73; p = 0.0523)	29 percent	36 percent	14 percent	22 percent
RT-1234–Treasure	1	*2*	3	4
Hider (187; p = 0.0036)	25 percent	22 percent	36 percent	18 percent
Seeker (84; p = 3E–05)	20 percent	18 percent	48 percent	14 percent
RT-1234–Mine	1	*2*	3	4
Hider (133; p = 6E-06)	18 percent	20 percent	44 percent	17 percent
Seeker (72; p = 0.149)	19 percent	25 percent	36 percent	19 percent
R-ABAA	A	*B*	A	A
Hider (50; p = 0.0186)	16 percent	18 percent	44 percent	22 percent
Seeker (64; p = 9E-07)	16 percent	19 percent	54 percent	11 percent

Notes: Sample sizes and p-values for significant differences from equilibrium in parentheses; salient labels in italics; order of presentation of locations to subjects as shown.

Crawford and Iriberri's Table 1

RTH took the main patterns in their data as evidence that their subjects did not think strategically:

- “The finding that both choosers and guessers selected the least salient alternative suggests little or no strategic thinking.”

- “In the competitive games, however, the players employed a naïve strategy (avoiding the endpoints), that is not guided by valid strategic reasoning. In particular, the hiders in this experiment either did not expect that the seekers too, will tend to avoid the endpoints, or else did not appreciate the strategic consequences of this expectation.”

But Crawford and Iriberri’s analysis suggests that RTH’s subjects were actually more than usually sophisticated (with many *L3*s and even *L4*s, although in most settings *L1*s and *L2*s are more common)—their thinking was plainly strategic, but just didn’t follow equilibrium logic.

Crawford and Iriberri’s analysis suggests that the Yushchenko quotation simply reflects the reasoning of an *L1* poisoner, or equivalently of an *L2* investigator reasoning about an *L1* poisoner.

RTH's results raise a couple of puzzles:

- On average hiders are as smart as seekers, so hiders tempted to hide in *central A* should realize that seekers will be just as tempted to look there. Why do hiders allow seekers to find them 32% of the time when they could hold it to 25% via the equilibrium mixed strategy?

- Further, why do seekers choose *central A* (or its analogs) even more often (46% in Table 3 below) than hiders (37%)?

Although the payoff structure of RTH's game is asymmetric, QRE ignores labeling and (logit or not) coincides with equilibrium in the game, and so neither helps to explain the asymmetry of choice distributions.

The role asymmetry in subjects' behavior and how it is linked to the game's payoffs points strongly in the direction of a level-*k*/CH model, and is a mystery from the viewpoint of other theories I am aware of.

n constructing such a level-*k* model, defining *L0* as uniform random would be unnatural, given the non-neutral framing of decisions and that *L0* describes others' instinctive responses.

It would also make *Lk* the same as *Equilibrium* for $k > 0$.

But a level-*k* model with a role-independent *L0* that probabilistically favors salient locations yields a simple explanation of RTH's results.

Assume that *L0* hiders and seekers both choose A, B, A, A with probabilities $p/2$, q, $1– p – q$, $p/2$ respectively, with $p > ½$ and $q > ¼$.

L0 favors both the end locations and the B location, equally for hiders and seekers, but the model lets the data decide which is more salient.

For behaviorally plausible type distributions (estimated 0% *L0*, 19% *L1*, 32% *L2*, 24% *L3*, 25% *L4*—almost hump-shaped), a level-*k* model gracefully explains the major patterns in RTH's data, the prevalence of *central A* for hiders and its even greater prevalence for seekers:

- Given *L0*'s attraction to salient locations, *L1* hiders choose *central A* to avoid *L0* seekers and *L1* seekers avoid *central A* searching for *L0* hiders (the data suggest that end locations are more salient than B).

- For similar reasons, *L2* hiders choose *central A* with probability between 0 and 1 (breaking payoff ties randomly) and *L2* seekers choose it with probability 1.

- *L3* hiders avoid *central A* and *L3* seekers choose it with probability between zero and one (breaking payoff ties randomly).

- *L4* hiders and seekers both avoid *central A*.

TABLE 2—TYPES' EXPECTED PAYOFFS AND CHOICE PROBABILITIES IN RTH'S GAMES WHEN $p > 1/2$ AND $q > 1/4$

Hider	Expected payoff	Choice probability	Expected payoff	Choice probability	Seeker	Expected payoff	Choice probability	Expected payoff	Choice probability
	$p < 2q$	$p < 2q$	$p > 2q$	$p > 2q$		$p < 2q$	$p < 2q$	$p > 2q$	$p > 2q$
L0 (Pr, r)					*L0* (Pr, r)				
A	–	$p/2$	–	$p/2$	A	–	$p/2$	–	$p/2$
B	–	q	–	q	B	–	q	–	q
A	–	$1-p-q$	–	$1-p-q$	A	–	$1-p-q$	–	$1-p-q$
A	–	$p/2$	–	$p/2$	A	–	$p/2$	–	$p/2$
L1 (Pr, s)					*L1* (Pr, s)				
A	$1 - p/2 < 3/4$	0	$1 - p/2 < 3/4$	0	A	$p/2 > 1/4$	0	$p/2 > 1/4$	1/2
B	$1 - q < 3/4$	0	$1 - q < 3/4$	0	B	$q > 1/4$	1	$q > 1/4$	0
A	$p + q > 3/4$	1	$p + q > 3/4$	1	A	$1-p-q < 1/4$	0	$1-p-q < 1/4$	0
A	$1 - p/2 < 3/4$	0	$1 - p/2 < 3/4$	0	A	$p/2 > 1/4$	0	$p/2 > 1/4$	1/2
L2 (Pr, t)					*L2* (Pr, t)				
A	1	1/3	1/2	0	A	0	0	0	0
B	0	0	1	1/2	B	0	0	0	0
A	1	1/3	1	1/2	A	1	1	1	1
A	1	1/3	1/2	0	A	0	0	0	0
L3 (Pr, u)					*L3* (Pr, u)				
A	1	1/3	1	1/3	A	1/3	1/3	0	0
B	1	1/3	1	1/3	B	0	0	1/2	1/2
A	0	0	0	0	A	1/3	1/3	1/2	1/2
A	1	1/3	1	1/3	A	1/3	1/3	0	0
L4 (Pr, v)					*L4* (Pr, v)				
A	2/3	0	1	1/2	A	1/3	1/3	1/3	1/3
B	1	1	1/2	0	B	1/3	1/3	1/3	1/3
A	2/3	0	1/2	0	A	0	0	0	0
A	2/3	0	1	1/2	A	1/3	1/3	1/3	1/3
Total	$p < 2q$		$p > 2q$		Total	$p < 2q$		$p > 2q$	
A	$rp/2+(1-\varepsilon)[t/3+u/3]$ $+(1-r)\varepsilon/4$		$rp/2+(1-\varepsilon)[u/3+w/2]$ $+(1-r)\varepsilon/4$		A	$rp/2+(1-\varepsilon)[u/3+w/3]$ $+(1-r)\varepsilon/4$		$rp/2+(1-\varepsilon)[s/2+w/3]$ $+(1-r)\varepsilon/4$	
B	$rq+(1-\varepsilon)[u/3+v]$ $+(1-r)\varepsilon/4$		$rq+(1-\varepsilon)[t/2+u/3]$ $+(1-r)\varepsilon/4$		B	$rq+(1-\varepsilon)[s+v/3]$ $+(1-r)\varepsilon/4$		$rq+(1-\varepsilon)[u/2+w/3]$ $+(1-r)\varepsilon/4$	
A	$r(1-p-q)+(1-\varepsilon)[s+t/3]$ $+(1-r)\varepsilon/4$		$r(1-p-q)+(1-\varepsilon)[s+t/2]$ $+(1-r)\varepsilon/4$		A	$r(1-p-q)+(1-\varepsilon)[t+u/3]$ $+(1-r)\varepsilon/4$		$r(1-p-q)+(1-\varepsilon)[t+u/2]$ $+(1-r)\varepsilon/4$	
A	$rp/2+(1-\varepsilon)[t/3+u/3]$ $+(1-r)\varepsilon/4$		$rp/2+(1-\varepsilon)[u/3+w/2]$ $+(1-r)\varepsilon/4$		A	$rp/2+(1-\varepsilon)[u/3+w/3]$ $+(1-r)\varepsilon/4$		$rp/2+(1-\varepsilon)[s/2+w/3]$ $+(1-r)\varepsilon/4$	

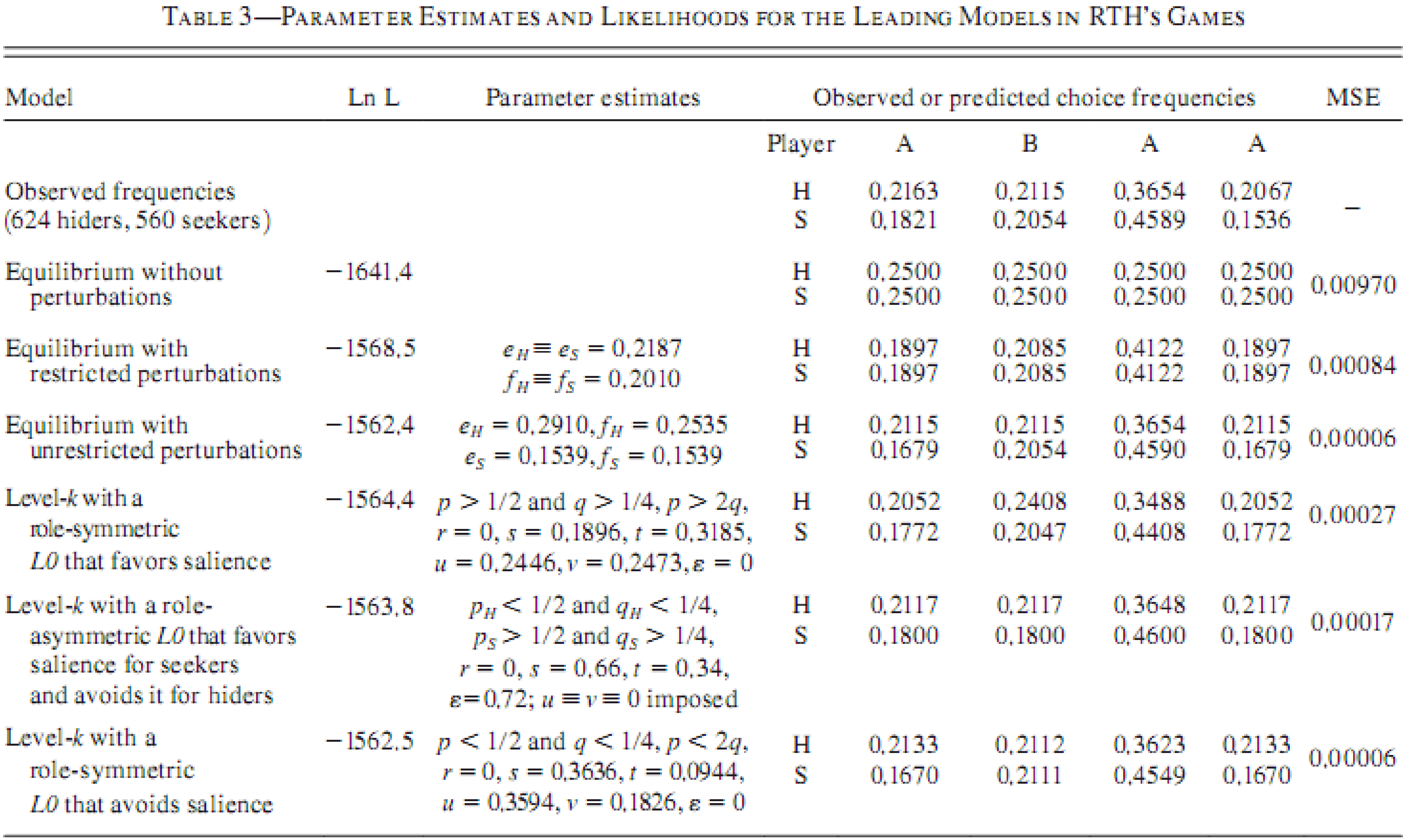


TABLE 3—PARAMETER ESTIMATES AND LIKELIHOODS FOR THE LEADING MODELS IN RTH'S GAMES

Model	Ln L	Parameter estimates	Observed or predicted choice frequencies					MSE
			Player	A	B	A	A	
Observed frequencies (624 hiders, 560 seekers)			H S	0,2163 0,1821	0,2115 0,2054	0,3654 0,4589	0,2067 0,1536	–
Equilibrium without perturbations	−1641,4		H S	0,2500 0,2500	0,2500 0,2500	0,2500 0,2500	0,2500 0,2500	0,00970
Equilibrium with restricted perturbations	−1568,5	$e_H \equiv e_S = 0,2187$ $f_H \equiv f_S = 0,2010$	H S	0,1897 0,1897	0,2085 0,2085	0,4122 0,4122	0,1897 0,1897	0,00084
Equilibrium with unrestricted perturbations	−1562,4	$e_H = 0,2910, f_H = 0,2535$ $e_S = 0,1539, f_S = 0,1539$	H S	0,2115 0,1679	0,2115 0,2054	0,3654 0,4590	0,2115 0,1679	0,00006
Level-*k* with a role-symmetric *L0* that favors salience	−1564,4	$p > 1/2$ and $q > 1/4, p > 2q,$ $r = 0, s = 0,1896, t = 0,3185,$ $u = 0,2446, v = 0,2473, \varepsilon = 0$	H S	0,2052 0,1772	0,2408 0,2047	0,3488 0,4408	0,2052 0,1772	0,00027
Level-*k* with a role-asymmetric *L0* that favors salience for seekers and avoids it for hiders	−1563,8	$p_H < 1/2$ and $q_H < 1/4,$ $p_S > 1/2$ and $q_S > 1/4,$ $r = 0, s = 0,66, t = 0,34,$ $\varepsilon = 0,72$; $u \equiv v \equiv 0$ imposed	H S	0,2117 0,1800	0,2117 0,1800	0,3648 0,4600	0,2117 0,1800	0,00017
Level-*k* with a role-symmetric *L0* that avoids salience	−1562,5	$p < 1/2$ and $q < 1/4, p < 2q,$ $r = 0, s = 0,3636, t = 0,0944,$ $u = 0,3594, v = 0,1826, \varepsilon = 0$	H S	0,2133 0,1670	0,2112 0,2111	0,3623 0,4549	0,2133 0,1670	0,00006

Crawford and Iriberri's Table 3

Note that only a heterogeneous population with substantial frequencies of *L2* and *L3* as well as *L1* (estimated 0% *L0*, 19% *L1*, 32% *L2*, 24% *L3*, 25% *L4*) can reproduce the aggregate patterns in the data.

(Even though there is a nonnegligible estimated frequency of *L4*s, they don’t really matter here because they never choose *central A* (Table 2 above), hence they are not implicated in the major aggregate patterns.

For the same reason, their frequency is not well identified in the estimation.)

For example, Crawford and Iriberri estimate (Table 3 above, row 5) that the salience of an end location is greater than the salience of the *B* ($p > 2q$).

Given this, a 50-50 mix of *L1*s and *L2*s in both player roles would imply (Table 2 above, right-most columns in each panel) 75% of hiders but only 50% of seekers choosing *central A*, in contrast to the 37% of hiders and 46% of seekers who did choose *central A*.

In Crawford and Iriberri's analysis of RTH's data, the role asymmetry in aggregate behavior follows naturally from the asymmetry of the game's payoff structure, via hiders' and seekers' asymmetric responses to *L0*'s *role-symmetric* choices.

Allowing *L0* to vary across roles as in Bacharach and Stahl (2000), although it yields a small improvement in fit (Table 3), would beg the question of why subjects' responses were so role-asymmetric.

The freedom to specify *L0* leaves room for doubts about overfitting and portability, the extent to which a model estimated from responses to one game can be extended to predict responses to different games.

Crawford and Iriberri tested for overfitting, confirming the level-*k* model.

A more challenging test regards portability.

Crawford and Iriberri tested for portability by using the leading alternative models, estimated from RTH's data, to "predict" subjects' initial responses in the two closest relatives of RTH's games in the literature:

● O'Neill's 1987 *PNAS* famous card-matching game, and

● Rapoport and Boebel's 1992 *GEB* closely related game.

These games both raise the same kinds of strategic issues as RTH's games, but with more complex patterns of wins and losses, different framing, and in the latter case five locations.

I focus here on Crawford and Iriberri's analysis of O'Neill's game.

In O’Neill’s card-matching game, players simultaneously and independently choose one of four cards: A, 2, 3, J.

One player, say the row player—but the game was presented to subjects as a story, not a matrix—wins if there is a match on J or a mismatch on A, 2, or 3; the other player wins in the other cases.

	A	**2**	**3**	**J**
A	1 0	0 1	0 1	1 0
2	0 1	1 0	0 1	1 0
3	0 1	0 1	1 0	1 0
J	1 0	1 0	1 0	0 1

O’Neill’s card-matching game

O’Neill’s game is like a hide-and-seek game, except that each player is a hider (h) for some locations and a seeker (s) for others.

A, 2, and 3 are strategically symmetric, and equilibrium (without payoff perturbations) has Pr{A} = Pr{2} = Pr{3} = 0.2, Pr{J} = 0.4.

	A (s)	2 (s)	3 (s)	J (h)
A (h)	1 0	0 1	0 1	1 0
2 (h)	0 1	1 0	0 1	1 0
3 (h)	0 1	0 1	1 0	1 0
J (s)	1 0	1 0	1 0	0 1

O’Neill’s card-matching game

The portability test directly addresses the issue of whether level-*k* models allow the modeler too much flexibility.

With regard to the flexibility of *L0*, first consider how to adapt our "psychological" specification of *L0* from RTH's to O'Neill's game.

"Anyone" should agree on the right kind of *L0*:

● A and J, "face" cards and end locations, are more salient than 2 and 3, but the specification should allow either A or J to be more salient.

That the RTH estimates suggested that their end locations are more salient than the *B* label does *not* dictate whether A or J is more salient, though it does reinforce that they are both more salient than 2 and 3.

This is a psychological issue, but because it is "only" a psychological issue, it is easy to gather evidence on it from different settings, and such evidence is more likely to yield convergence than if it were partly a strategic issue.

Further, because all that matters about *L0* is what it makes *L1*s do in each role, the remaining freedom to choose *L0* allows only two models.

With regard to the flexibility of the type frequencies, empirically plausible frequencies often imply severe limits on what decision patterns a level-*k* model can generate.

Discussions of O'Neill's data (which we did not have before we carried out the analysis), had been dominated by an "Ace effect":

Aggregated over all 105 rounds, row and column players played A with frequencies 22.0% and 22.6%, significantly above the equilibrium 20%.

O'Neill speculated that this was because "…players were attracted by the powerful connotations of an Ace".

Yet no behaviorally plausible level-*k* model can make a row player play A more than the equilibrium 20%.

Crawford and Iriberri's online appendix, Tables A3 and A4, show that, excluding *L0*s as normally having 0 estimated frequencies and restricting attention to row players (Player 1s), when A is more salient (3*j* – *a* < 1) only *L4* chooses A, with probability at most 1/3; and when A is less salient (3*j* – *a* > 1) only *L3* chooses A, with probability at most 1/3.

This is logically possible, but in the first case it would require a population of 60% or more *L4*s, and in the second case it would require 60% or more *L3*s.)

Thus, despite the flexibility of the estimated type distribution, the level-*k* model's structure and the principles that guide the specification of *L0* imply that row players play A less than the equilibrium 20%.

Speculating that O'Neill's subjects' *initial* responses must not have had an Ace effect, Crawford and Iriberri got the data and found that there was in fact no Ace effect for initial responses.

Instead there was a Joker effect, a full order of magnitude stronger, in which row players played J 56% of the time and column players played it 64% of the time.

Unlike the putative Ace effect, the Joker effect and the other observed frequencies *can* be gracefully explained by a level-*k* model with an *L0* that probabilistically favors the salient A and J cards, in the spirit of Crawford and Iriberri's analysis of RTH's data.

Crawford and Iriberri's analysis traces the portability of the level-*k* model to the fact that *L0* is psychological rather than strategic, and that it is based on simple and universal intuition and evidence.

If *L0* were strategic, it would interact with the strategic structure in new ways in each new game, and seldom could one extrapolate a specification from one game to another.

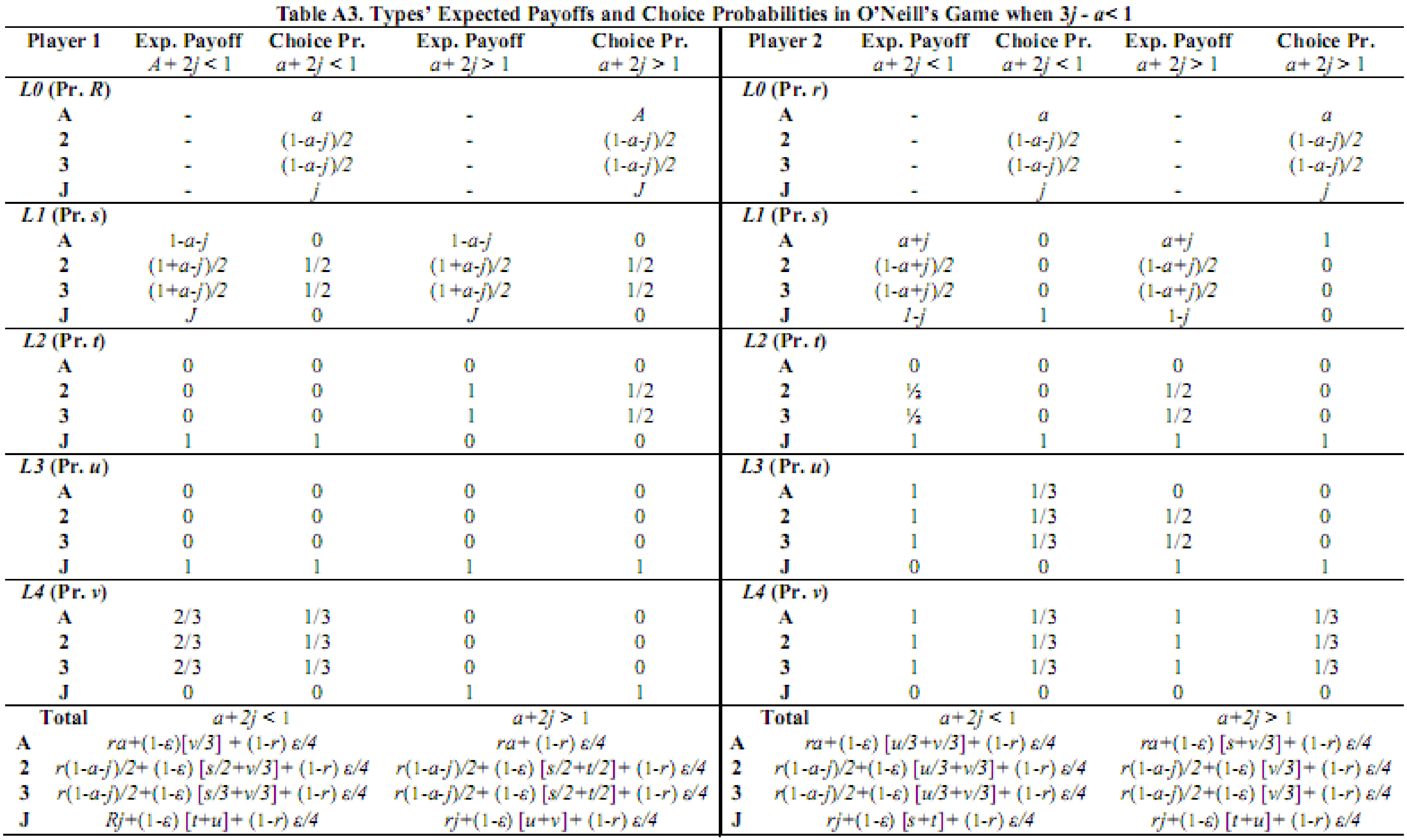


Table A3. Types' Expected Payoffs and Choice Probabilities in O'Neill's Game when $3j - a < 1$

Player 1	Exp. Payoff	Choice Pr.	Exp. Payoff	Choice Pr.	Player 2	Exp. Payoff	Choice Pr.	Exp. Payoff	Choice Pr.
	$A+2j<1$	$a+2j<1$	$a+2j>1$	$a+2j>1$		$a+2j<1$	$a+2j<1$	$a+2j>1$	$a+2j>1$
L0 **(Pr. *R*)**					***L0*** **(Pr. *r*)**				
A	-	a	-	A	**A**	-	a	-	a
2	-	$(1-a-j)/2$	-	$(1-a-j)/2$	**2**	-	$(1-a-j)/2$	-	$(1-a-j)/2$
3	-	$(1-a-j)/2$	-	$(1-a-j)/2$	**3**	-	$(1-a-j)/2$	-	$(1-a-j)/2$
J	-	j	-	J	**J**	-	j	-	j
L1 **(Pr. *s*)**					***L1*** **(Pr. *s*)**				
A	$1-a-j$	0	$1-a-j$	0	**A**	$a+j$	0	$a+j$	1
2	$(1+a-j)/2$	1/2	$(1+a-j)/2$	1/2	**2**	$(1-a+j)/2$	0	$(1-a+j)/2$	0
3	$(1+a-j)/2$	1/2	$(1+a-j)/2$	1/2	**3**	$(1-a+j)/2$	0	$(1-a+j)/2$	0
J	J	0	J	0	**J**	$1-j$	1	$1-j$	0
L2 **(Pr. *t*)**					***L2*** **(Pr. *t*)**				
A	0	0	0	0	**A**	0	0	0	0
2	0	0	1	1/2	**2**	½	0	1/2	0
3	0	0	1	1/2	**3**	½	0	1/2	0
J	1	1	0	0	**J**	1	1	1	1
L3 **(Pr. *u*)**					***L3*** **(Pr. *u*)**				
A	0	0	0	0	**A**	1	1/3	0	0
2	0	0	0	0	**2**	1	1/3	1/2	0
3	0	0	0	0	**3**	1	1/3	1/2	0
J	1	1	1	1	**J**	0	0	1	1
L4 **(Pr. *v*)**					***L4*** **(Pr. *v*)**				
A	2/3	1/3	0	0	**A**	1	1/3	1	1/3
2	2/3	1/3	0	0	**2**	1	1/3	1	1/3
3	2/3	1/3	0	0	**3**	1	1/3	1	1/3
J	0	0	1	1	**J**	0	0	0	0
Total	$a+2j<1$		$a+2j>1$		**Total**	$a+2j<1$		$a+2j>1$	
A	$ra+(1-\varepsilon)[v/3] + (1-r)\ \varepsilon/4$		$ra+ (1-r)\ \varepsilon/4$		**A**	$ra+(1-\varepsilon)\ [u/3+v/3]+ (1-r)\ \varepsilon/4$		$ra+(1-\varepsilon)\ [s+v/3]+ (1-r)\ \varepsilon/4$	
2	$r(1-a-j)/2+ (1-\varepsilon)\ [s/2+v/3]+ (1-r)\ \varepsilon/4$		$r(1-a-j)/2+ (1-\varepsilon)\ [s/2+t/2]+ (1-r)\ \varepsilon/4$		**2**	$r(1-a-j)/2+(1-\varepsilon)\ [u/3+v/3]+ (1-r)\ \varepsilon/4$		$r(1-a-j)/2+(1-\varepsilon)\ [v/3]+ (1-r)\ \varepsilon/4$	
3	$r(1-a-j)/2+(1-\varepsilon)\ [s/3+v/3]+ (1-r)\ \varepsilon/4$		$r(1-a-j)/2+ (1-\varepsilon)\ [s/2+t/2]+ (1-r)\ \varepsilon/4$		**3**	$r(1-a-j)/2+(1-\varepsilon)\ [u/3+v/3]+ (1-r)\ \varepsilon/4$		$r(1-a-j)/2+(1-\varepsilon)\ [v/3]+ (1-r)\ \varepsilon/4$	
J	$Rj+(1-\varepsilon)\ [t+u]+ (1-r)\ \varepsilon/4$		$rj+(1-\varepsilon)\ [u+v]+ (1-r)\ \varepsilon/4$		**J**	$rj+(1-\varepsilon)\ [s+t]+ (1-r)\ \varepsilon/4$		$rj+(1-\varepsilon)\ [t+u]+ (1-r)\ \varepsilon/4$	

Table A4. Types' Expected Payoffs and Choice Probabilities in O'Neill's Game when $3j - a > 1$

Player 1	Exp. Payoff	Choice Pr.	Player 2	Exp. Payoff	Choice Pr.
***L0* (Pr. *R*)**			***L0* (Pr. *r*)**		
A	-	a	**A**	-	a
2	-	$(1-a-j)/2$	**2**	-	$(1-a-j)/2$
3	-	$(1-a-j)/2$	**3**	-	$(1-a-j)/2$
J	-	j	**J**	-	j
***L1* (Pr. *S*)**			***L1* (Pr. *s*)**		
A	$1-a-j$	0	**A**	$a+j$	1
2	$(1+a-j)/2$	*0*	**2**	$(1-a+j)/2$	0
3	$(1+a-j)/2$	*0*	**3**	$(1-a+j)/2$	0
J	j	1	**J**	$1-j$	0
***L2* (Pr. *T*)**			***L2* (Pr. *t*)**		
A	0	0	**A**	1	1/3
2	1	1/2	**2**	1	1/3
3	1	1/2	**3**	1	1/3
J	0	0	**J**	0	0
***L3* (Pr. *U*)**			***L3* (Pr. *u*)**		
A	2/3	1/3	**A**	0	0
2	2/3	1/3	**2**	1/2	0
3	2/3	1/3	**3**	1/2	0
J	0	0	**J**	1	1
***L4* (Pr. *V*)**			***L4* (Pr. *v*)**		
A	0	0	**A**	1/3	0
2	0	0	**2**	1/3	0
3	0	0	**3**	1/3	0
J	1	1	**J**	1	1
Total			**Total**		
A	$Ra+(1-\varepsilon)[u/3]+(1-r)\,\varepsilon/4$		**A**	$ra+(1-\varepsilon)[s+t/3]+(1-r)\,\varepsilon/4$	
2	$r(1-a-j)/2+(1-\varepsilon)[t/2+u/3]+(1-r)\,\varepsilon/4$		**2**	$r(1-a-j)/2+(1-\varepsilon)[t/3]+(1-r)\,\varepsilon/4$	
3	$R(1-a-j)/2+(1-\varepsilon)[t/2+u/3]+(1-r)\,\varepsilon/4$		**3**	$r(1-a-j)/2+(1-\varepsilon)[t/3]+(1-r)\,\varepsilon/4$	
J	$Rj+(1-\varepsilon)[s+v]+(1-r)\,\varepsilon/4$		**J**	$rj+(1-\varepsilon)[u+v]+(1-r)\,\varepsilon/4$	

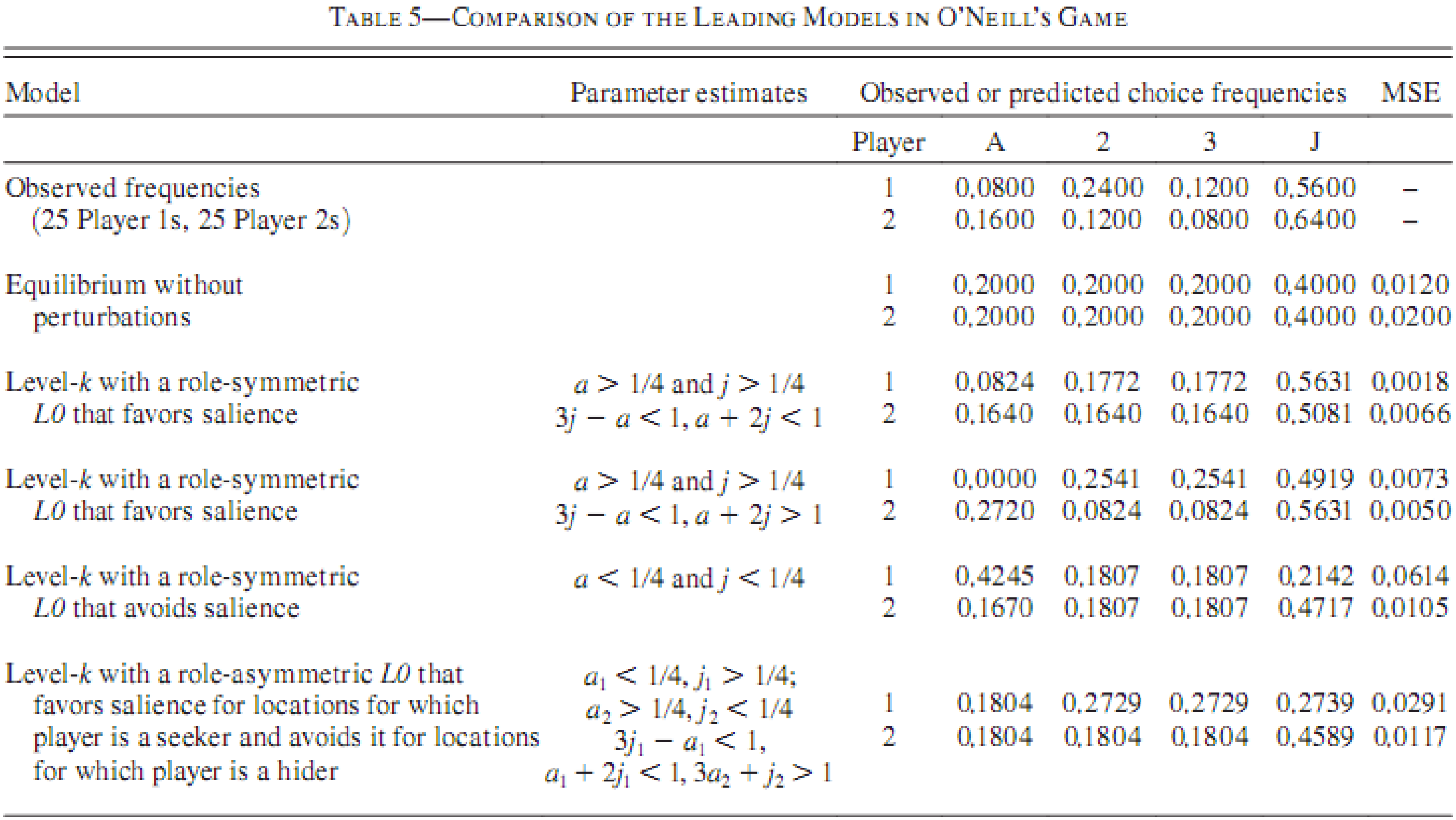


Table 5—Comparison of the Leading Models in O'Neill's Game

Model	Parameter estimates	Observed or predicted choice frequencies					MSE
		Player	A	2	3	J	
Observed frequencies (25 Player 1s, 25 Player 2s)		1	0,0800	0,2400	0,1200	0,5600	–
		2	0,1600	0,1200	0,0800	0,6400	–
Equilibrium without perturbations		1	0,2000	0,2000	0,2000	0,4000	0,0120
		2	0,2000	0,2000	0,2000	0,4000	0,0200
Level-*k* with a role-symmetric *L0* that favors salience	$a > 1/4$ and $j > 1/4$ $3j - a < 1, a + 2j < 1$	1	0,0824	0,1772	0,1772	0,5631	0,0018
		2	0,1640	0,1640	0,1640	0,5081	0,0066
Level-*k* with a role-symmetric *L0* that favors salience	$a > 1/4$ and $j > 1/4$ $3j - a < 1, a + 2j > 1$	1	0,0000	0,2541	0,2541	0,4919	0,0073
		2	0,2720	0,0824	0,0824	0,5631	0,0050
Level-*k* with a role-symmetric *L0* that avoids salience	$a < 1/4$ and $j < 1/4$	1	0,4245	0,1807	0,1807	0,2142	0,0614
		2	0,1670	0,1807	0,1807	0,4717	0,0105
Level-*k* with a role-asymmetric *L0* that favors salience for locations for which player is a seeker and avoids it for locations for which player is a hider	$a_1 < 1/4, j_1 > 1/4;$ $a_2 > 1/4, j_2 < 1/4$ $3j_1 - a_1 < 1,$ $a_1 + 2j_1 < 1, 3a_2 + j_2 > 1$	1	0,1804	0,2729	0,2729	0,2739	0,0291
		2	0,1804	0,1804	0,1804	0,4589	0,0117

Crawford and Iriberri's Table 5

Application: Huarangdao and D-Day: Communication of Intentions in Outguessing Games

General Kongming: “Have you forgotten the tactic of ‘letting weak points look weak and strong points look strong’?”

General 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.’”

—Luo Guanzhong’s historical novel, *Three Kingdoms*

The quotations refer to a two-person outguessing game with complete information and one-sided communication of intentions via cheap talk.

In the story, set around 200 A.D., fleeing general Cao Cao chose between two escape routes, the easier Main Road and the awful Huarong Road, trying to avoid capture by pursuing General Kongming (http://en.wikipedia.org/wiki/Battle_of_Red_Cliffs).

Kongming (the sender in this example) waited in ambush along the Huarong Road and set campfires there, thus sending a deceptively truthful message.

Cao Cao (the receiver), misjudging Kongming’s communication strategy, inverted the truthful message and was caught by Kongming (but was later released).

Consider a simple model of the underlying game (without communication):

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

Huarongdao

(Remarkably, Huarongdao has the same payoffs as *Far Pavilions* Escape.)

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

- But both Cao Cao and Kongming gain 1 by taking the Main Road, whether or not Cao Cao is captured: It’s important to be comfortable, even if (especially if?) if you think you’re about to die.

The key issues here are how Kongming should choose his message and how Cao Cao—knowing Kongming is choosing strategically to anticipate Cao Cao's interpretation—should interpret Kongming's message.

In real settings like this, a receiver's thinking often assigns a prominent role to the literal meanings of messages, without necessarily taking them at face value; and a sender's thinking takes this into account.

But a standard equilibrium analysis precludes a role for the literal meanings of payoff-irrelevant messages (Crawford and Sobel 1982 *Econometrica*; see however Farrell's 1993 *GEB* neologism-proofness refinement, which depends on meanings).

Moreover, there is no equilibrium (refined or not) in a zero-sum (or this) two-person game in which cheap talk conveys information or the receiver responds to the message.

In such an equilibrium, if there was information in the sender’s message that the receiver found it optimal to respond to, the receiver’s response would help him and so hurt the sender, who would then prefer to make his message uninformative.

Thus communication can have no effect in any equilibrium, and as a result the underlying game must be played according to its unique mixed-strategy equilibrium, as if there were no communication phase.

Yet intuition suggests that in many such situations:

- The sender’s message and action are part of a single, integrated strategy.

- The sender tries to anticipate which message will fool the receiver and chooses it nonrandomly.

- The sender’s action differs from what he would have chosen with no opportunity to send a message.

Huarongdao is only one datapoint (and possibly fictional!). But there's another example in which the same thing may have happened.

Consider the Allies' choice of where to invade Europe on D-Day (6 June 1944), the motivating example of Crawford (2003 *AER*).

The underlying game can also be modeled as an outguessing game:

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

- Attacking an undefended Calais is better for the Allies than attacking an undefended Normandy, 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 them on average.

Now imagine that D-Day is preceded by a message from the Allies to the Germans regarding their intentions about where to attack, as in Operation Fortitude South (http://en.wikipedia.org/wiki/Operation_Fortitude).

Imagine further that the message is (approximately!) cheap talk.

A “Tank” from Operation Fortitude

In what sense did the same thing happen in Huarongdao and D-Day?

In each case the deception succeeded, but the sender won in the less beneficial of the two possible ways to win.

Kongming's message was literally truthful—he lit fires on the Huarong Road and ambushed Cao Cao there—but Cao Cao was fooled because he misread Kongming’s message strategy and inverted the message.

In D-Day the message was literally deceptive but the Germans were fooled because they “believed” it (either because they were credulous or because they inverted the message one too many times).

The sender’s and receiver’s message strategies and beliefs were different, but the outcome in the underlying game was the same: The sender won, but in the less beneficial of the two possible ways.

The quotations tell us why Cao Cao was fooled by Kongming’s message: with a truthful *L0*, his rationale resembles *L1* thinking, while Kongming’s resembles *L2* thinking.

But the quotations don’t tell us why in each case the sender won only in the less beneficial way.

General Kongming: “Have you forgotten the tactic of ‘letting weak points look weak and strong points look strong’?”

General 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.’”

To restate the puzzle more concretely, for both D-Day and Huarongdao:

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

- If the sender expected his message to fool the receiver, why didn't he reverse it and fool the receiver in the way that would have allowed him to win in the *more* beneficial way?

  (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?)

A level-*k* analysis suggests that it was more than a coincidence that the same thing happened in both cases.

Although *Sophisticated* subjects are rare in laboratory experiments, they may be more common in the field; and it is interesting to see whether a plausible model allows deception between *Sophisticated* players.

Accordingly, let Allies' and Germans' types be drawn from separate distributions, each including both level-*k* or *Mortal* types (in honor of Puck, in *A Midsummer Night's Dream*, Act 3: "Lord, what fools these mortals be!") and a fully strategically rational or *Sophisticated*, type.

Mortal types use step-by-step procedures that generically determine unique pure strategies, and avoid simultaneous determination of the kind used to define equilibrium; recall the Selten (1998 *EER*) quote above.

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

How should *L0* be adapted to an extensive-form game with communication?

Here a uniform random *L0* seems quite unnatural. For sender or receiver, the instinctive reaction to a message in a language one understands is surely to focus on its literal meaning, even if one ends up either lying or not taking the message at face value.

The level-*k* model therefore anchors *Mortal* types' messages and responses on *L0*s based on truthfulness for senders and credulity for receivers, just as in the informal literature on deception.

(The literature has not yet converged on whether *L0* receivers should be defined as credulous or uniform random—compare Ellingsen and Östling (2010 *AER*)—but the distinction is partly semantic because truthful *L0* senders imply that *L1* receivers are also credulous.)

Mortal Allied types' simplified models of other players make *L1* or higher *Mortal* Allied types always expect to fool the Germans, either by lying (like the Allies) or by telling the truth (like Kongming).

Given this, all *L1* or higher *Mortal* Allied types send a message they expect to make the Germans think they will attack Normandy, and then attack Calais.

If we knew the Allies and Germans were *Mortal*, we could now derive the model's implications from an estimate of the type frequencies of *Mortal* Allies who tell the truth or lie, and of *Mortal* Germans who believe or invert the Allies' message.

But the analysis must also take into account the possibility of *Sophisticated* Allies and Germans, who know everything about the game, including the distribution of *Mortal* types, and play an equilibrium in the resulting game.

To take into account the possibility of *Sophisticated* Allies and Germans, note that *Mortal* players' strategies are determined independently of each other's and *Sophisticated* players' strategies, and so can be treated as exogenous (even though they affect other players' payoffs).

Plug in the distributions of *Mortal* Allies' and Germans' independently determined behaviors to obtain a "reduced game" between *Sophisticated* Allies and *Sophisticated* 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, ostensibly about his intentions, is in fact read by a *Sophisticated* receiver as a signal of the sender's type.

Thus, the possibility of *Mortal* players completely changes the character of the game between *Sophisticated* players, which is what gives the model the ability to explain the effectiveness of communication in a zero-sum game and the possibility of deception among *Sophisticated* players.

The equilibria of the reduced game are determined by the population frequencies of *Mortal* and *Sophisticated* senders and receivers. There are two leading cases, with different implications:

When *Sophisticated* Allies and Germans are common—not behaviorally plausible—the reduced game has a mixed-strategy 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-strategy equilibrium, in which *Sophisticated* Allies can predict *Sophisticated* Germans' decisions, and vice versa.

In the latter, pure-strategy equilibrium, *Sophisticated* Germans always defend Calais (because they know that *Mortal* Allies, who predominate when *Sophisticated* Allies are rare, will always attack Calais).

Sophisticated Allies cannot fool *Sophisticated* Germans, and so send the message that fools the most *Mortal* Germans (lying if more *Mortal* Germans believe than invert: no pure message can fool both kinds, or any *Sophisticated* Germans), and attack at the more profitable location.

Thus there is sometimes a pure-strategy “Fortitude” equilibrium in which *Sophisticated* Allies attack Normandy while *Sophisticated* Germans defend Calais.

But there is never a pure-strategy “reverse-Fortitude” equilibrium in which *Sophisticated* Allies attack Calais while *Sophisticated* Germans defend Normandy, even though that would be more profitable for Allies.

For, in such an equilibrium, any (pure) deviation from *Sophisticated* Allies’ equilibrium message would “prove” to *Sophisticated* Germans that the Allies were *Mortal*, making it optimal for them to defend Calais.

If *Sophisticated* Allies attack Calais in the equilibrium, the conclusion is immediate.

If, instead, *Sophisticated* Allies attack Normandy in the equilibrium while *Sophisticated* Germans defend Normandy, *Sophisticated* Allies’ message fools only the most likely kind of *Mortal* German (believers or inverters), with payoff gain of their frequency times the payoff of attacking an undefended Normandy.

But reversing the message and attack location would still fool the most likely kind of *Mortal* German, but now with payoff gain of their frequency times the higher payoff of attacking undefended Calais, a contradiction.

In that sense, the model explains why *Sophisticated* Allies don't attack Calais (or Kongming did not light campfires and ambush on Main Road).

In the pure-strategy equilibrium that exists when *Sophisticated* Allies and Germans are rare, the Allies' message and action are part of a single, integrated strategy; and the probability of attacking Normandy is much higher than if no communication 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 way.

Thus for plausible parameter values, with no unexplained difference in the sophistication of Allies and Germans, the model explains why *Sophisticated* Germans might allow themselves to be "fooled" by a costless message from a *Sophisticated* enemy: It is an unavoidable cost of exploiting mistakes by *Mortal* enemies, who are much more common.

Nonetheless, *Sophisticated* players in either role do strictly better than their *Mortal* counterparts.

Their advantage comes from the ability to avoid being fooled and/or to choose which *Mortal* type(s) to fool.

In the mixed-strategy equilibrium that prevails when *Sophisticated* Allies and Germans are common, *Sophisticated* players' equilibrium mixed strategies offset each other's gains from fooling *Mortal* Receivers, and in each role *Sophisticated* and *Mortal* players have equal expected payoffs.

This suggests that in an adaptive analysis of the dynamics of the type distribution, as in Conlisk (2001 *AER*), the frequencies of *Sophisticated* types will grow until the population is in the region of mixed-strategy equilibria in which types' expected payoffs are equal.

Thus *Sophisticated* and *Mortal* players can coexist in long-run equilibrium.

Application: Alphonse and Gaston: Communication of Intentions in Coordination Games

YOU FIRST, MY DEAR GASTON!
AFTER YOU, MY DEAR ALPHONSE!
F. Opper
COPYRIGHT, 1906, BY AMERICAN JOURNAL EXAMINER.
"YOU FIRST, MY DEAR."

"After you, Alphonse." "No, you first, my dear Gaston!"

—Frederick B. Opper's comic strip, *Alphonse and Gaston* (http://en.wikipedia.org/wiki/Alphonse_and_Gaston)

"What we got here…is a failure to communicate."

—Paul Newman as the title character in *Cool Hand Luke* (http://www.imdb.com/title/tt0061512/)

If level-*k* models allow preplay communication of intentions to affect the outcomes of zero-sum games, it should come as no surprise that they also allow effective communication in coordination games.

Ellingsen and Östling (2010 *AER*) and Crawford (2007), not discussed in detail here, adapt Crawford's (2003 *AER*) approach to study different aspects of preplay communication of intentions in coordination and other games.

Ellingsen and Östling use a level-*k* model to study the effectiveness of a single round of one- or two-sided preplay communication in games where communication of intentions plays various roles.

Crawford uses a level-*k* model to study the effectiveness of one- or multi-round two-sided communication in games like Battle of the Sexes, building on Farrell's 1987 *RAND J* and Rabin's 1994 *JET* analyses.

In each case the power of the analysis stems from the use of a model that does not assume equilibrium, which is question-begging in this context; but which imposes a realistic structure less agnostic than rationalizability.

Application: October Surprise: Communication of Private Information in Outguessing Games

"…The news that day was the so-called 'October Surprise' broadcast by bin Laden. He hadn't shown himself in nearly a year, but now, four days before the [2004 presidential] election, his spectral presence echoed into every American home. It was a surprisingly complete statement by the al Qaeda leader about his motivations, his actions, and his view of the current American landscape. He praised Allah and, through most of the eighteen minutes, attacked Bush,…. At the end, he managed to be dismissive of Kerry, but it was an afterthought in his 'anyone but Bush' treatise…. Inside the CIA…the analysis moved on a different track. They had spent years, as had a similar bin Laden unit at FBI, parsing each expressed word of the al Qaeda leader…. What they'd learned over nearly a decade is that bin Laden speaks only for strategic reasons…. Today's conclusion: bin Laden's message was clearly designed to help the President's reelection."

—Suskind, *The One Percent Doctrine*, 2006, pp. 335-6 (quoted in Jazayerli 2008 http://www.fivethirtyeight.com/2008/10/guest-column-will-bin-laden-strike.html).

October Surprise

The quotation refers to a zero-sum two-person game with incomplete information and one-sided preplay communication of private information via cheap talk.

Only bin Laden knows which candidate he wants; and, talk being cheap, he will say what it takes to help his candidate win.

A representative American voter knows only that he wants whichever candidate bin Laden doesn’t want.

The key issues are how bin Laden should relate his statement to what he really wants and how the American should interpret bin Laden’s statement, knowing that bin Laden is choosing the message strategically.

Once again, the literal meanings of messages are likely to play a prominent role in applications, but equilibrium analysis precludes such a role.

There is again no equilibrium in which cheap talk conveys information, or in which the receiver responds to the sender’s message.

Consider, however, a level-*k* model in which *L0* is anchored on truthfulness for the sender (bin Laden) and credulity for the receiver (American voter). (Or one could derive credulity for an *L1* receiver.)

An *L0* or *L1* American believes bin Laden’s message, and therefore votes for whichever candidate bin Laden attacks.

An *L0* bin Laden who wants Bush to win attacks Kerry, but an *L1* (*L2*) bin Laden who wants Bush to win attacks Bush to induce *L0* (*L1*) Americans to vote for Bush.

Given bin Laden’s choice, an *L0* or *L1* American then votes for Bush, but an *L2* American votes for Kerry.

Bin Laden chooses his message to fool the most prevalent kind of American—believer or inverter—as in Crawford’s (2003 *AER*) analysis.

An *L2* bin Laden believes Americans are *L1*, so believes that “reverse psychology” will be effective.

Experimental Evidence: Wang, Spezio, and Camerer (2010 *AER*)

I now discuss some experimental evidence on communication of private information in discretized versions of Crawford and Sobel's (1982 *Econometrica*) sender-receiver games.

Sender observes state S = 1, 2, 3, 4, or 5, sends message M = 1, 2, 3, 4, or 5. Receiver observes message, chooses action A = 1, 2, 3, 4, or 5.

The Receiver's choice of A determines the welfare of both:

● The Receiver's ideal outcome is A = S.

● The Sender's ideal outcome is A = S + b (or 5, if S + B > 5).

The Receiver's von Neumann-Morgenstern utility function is $110 – 20|S – A|^{1.4}$, and the Sender's is $110 – 20|S + b – A|^{1.4}$.

The difference in preferences varied across treatments: b = 0, 1, or 2.

Crawford and Sobel’s theoretical analysis characterized the possible equilibrium relationships between Sender’s observed S and Receiver’s choice of A, which determine the informativeness of communication.

They showed, for models with continuous state and action spaces that generalize Wang et al.’s examples (except for discreteness), that all equilibria are “partition equilibria”, in which the Sender partitions the set of states into contiguous groups and tells the Receiver, in effect, only which group his observation lies in.

For any difference in Sender’s and Receiver’s preferences (b), there is a range of equilibria, from a “babbling” equilibrium with one partition element to more informative equilibria that exist when b is small enough.

Under reasonable assumptions there is a “most informative” equilibrium, which has the most partition elements and gives the Receiver the highest ex ante (before the Sender observes the state) expected payoff.

As the preference difference decreases, the amount of information transmitted in the most informative equilibrium increases (measured by the correlation between S and A or the Receiver’s expected payoff).

The unambiguous part of Crawford and Sobel’s characterization of equilibrium concerns the possible relationships between S and A.

Because messages have no direct effect on payoffs (“cheap talk”), there is nothing to tie down their meanings in equilibrium.

As a result, any equilibrium relationship between S and A can be supported by any sufficiently rich language, with the meanings of messages determined by players’ equilibrium beliefs.

Behaviorally, however, in experiments like Wang et al.’s with a clear correspondence between state and message—S = 1, 2, 3, 4, or 5 and M = 1, 2, 3, 4, or 5—or where communication is in a common natural language, interpretations of messages are dictated by literal meanings.

Thus messages are always understood—even if not always believed.

Wang et al.’s data analysis therefore fixes the meanings of Sender subjects’ messages at their literal values.

Even with this restriction, when b = 0 or 1 in their design (Sender’s and Receiver’s preferences are close enough) there are multiple equilibria.

Wang et al.’s analysis then focuses on the “most informative” equilibrium.

When b = 0, the most informative equilibrium has M = S and A = S: perfect truth-telling, credulity, and perfect information transmission, as is intuitively plausible with identical preferences.

When b = 2, the most informative equilibrium has Senders sending a completely uninformative message M = {1, 2, 3, 4, 5} for any value of S; and Receivers ignoring it, hence choosing A = 3, which is optimal given their prior beliefs, for any value of M.

(A babbling equilibrium also exists when b = 0 or 1, but then it is not the most informative equilibrium.)

When b = 1, the most informative equilibrium has Senders sending M = 1 when S = 1 but M = {2, 3, 4, 5} when S = 2, 3, 4, or 5; and Receivers choosing A = 1 when M = 1 and A = 3 or 4 when M = {2, 3, 4, 5}.

(The Sender's message M = {2, 3, 4, 5} is the simplest way to implement the intentional vagueness of this partition equilibrium. Another way would be for the Sender to randomize M uniformly on {2, 3, 4, 5} when S = 1.)

Thus, when b = 1 the difference in preferences causes noisy information transmission even in the most informative equilibrium.

Importantly, however, in Crawford and Sobel's equilibrium analysis the Receiver's beliefs on hearing the Sender's message M are necessarily an unbiased—though noisy—estimate of S:

In equilibrium there can be no lying or deception as often occurs in real communication, only intentional vagueness (which also occurs).

Turning to Wang et al.’s experimental results, when b = 0 Senders almost always set M = S and Receivers almost always set A = M:

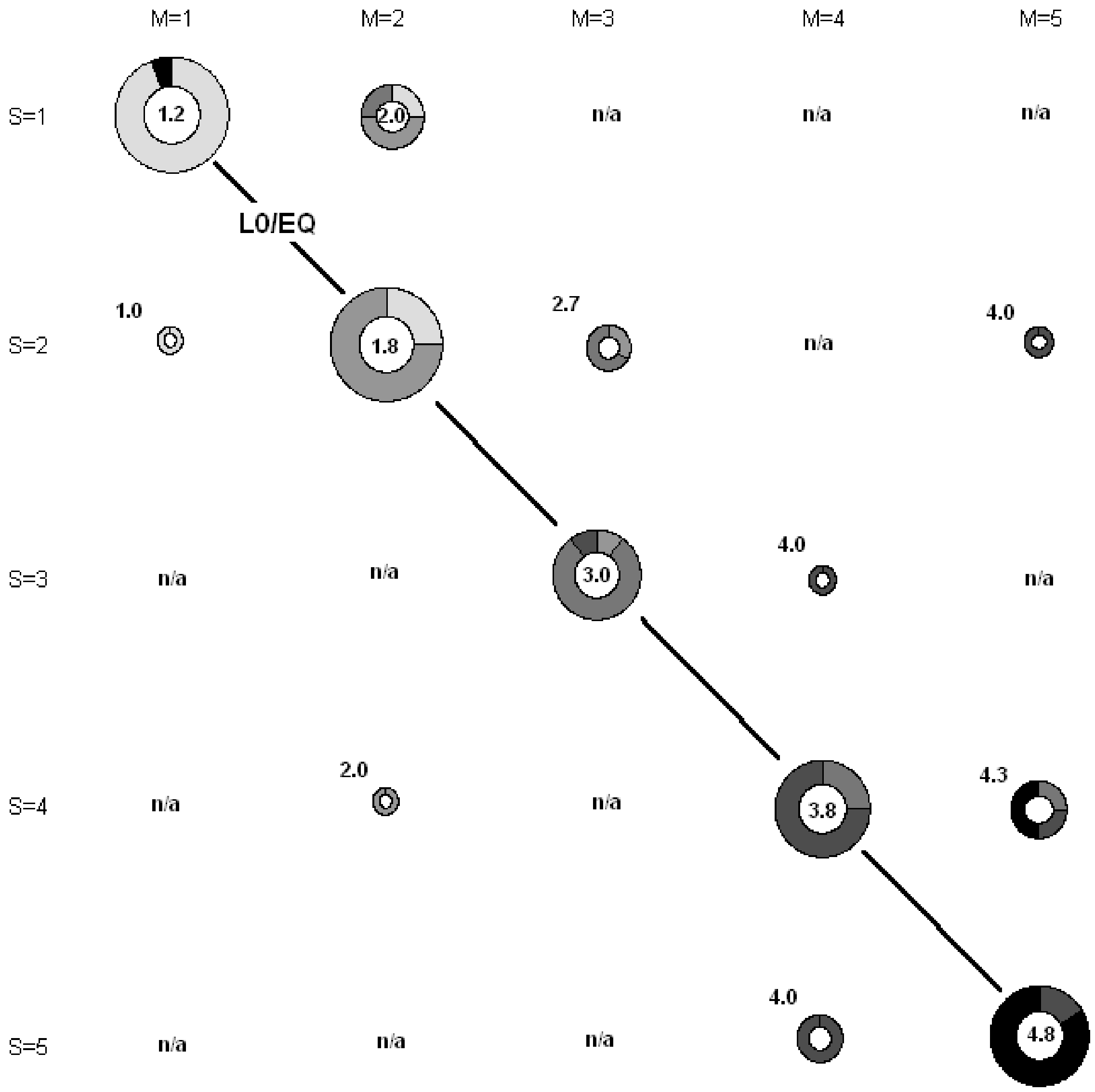
The result is near the perfect information transmission predicted by the most informative equilibrium.

Figure 1 (next slide) shows the Sender’s message frequencies and the Receiver’s action frequencies as functions of the observed state S when b = 0.

(A circle’s size shows the Sender’s message frequencies.

A circle’s darkness and the numbers inside show the Receiver’s action frequencies.)

Figure 1: Raw Data Pie Charts (b=0)   (Hidden Bias-Stranger)



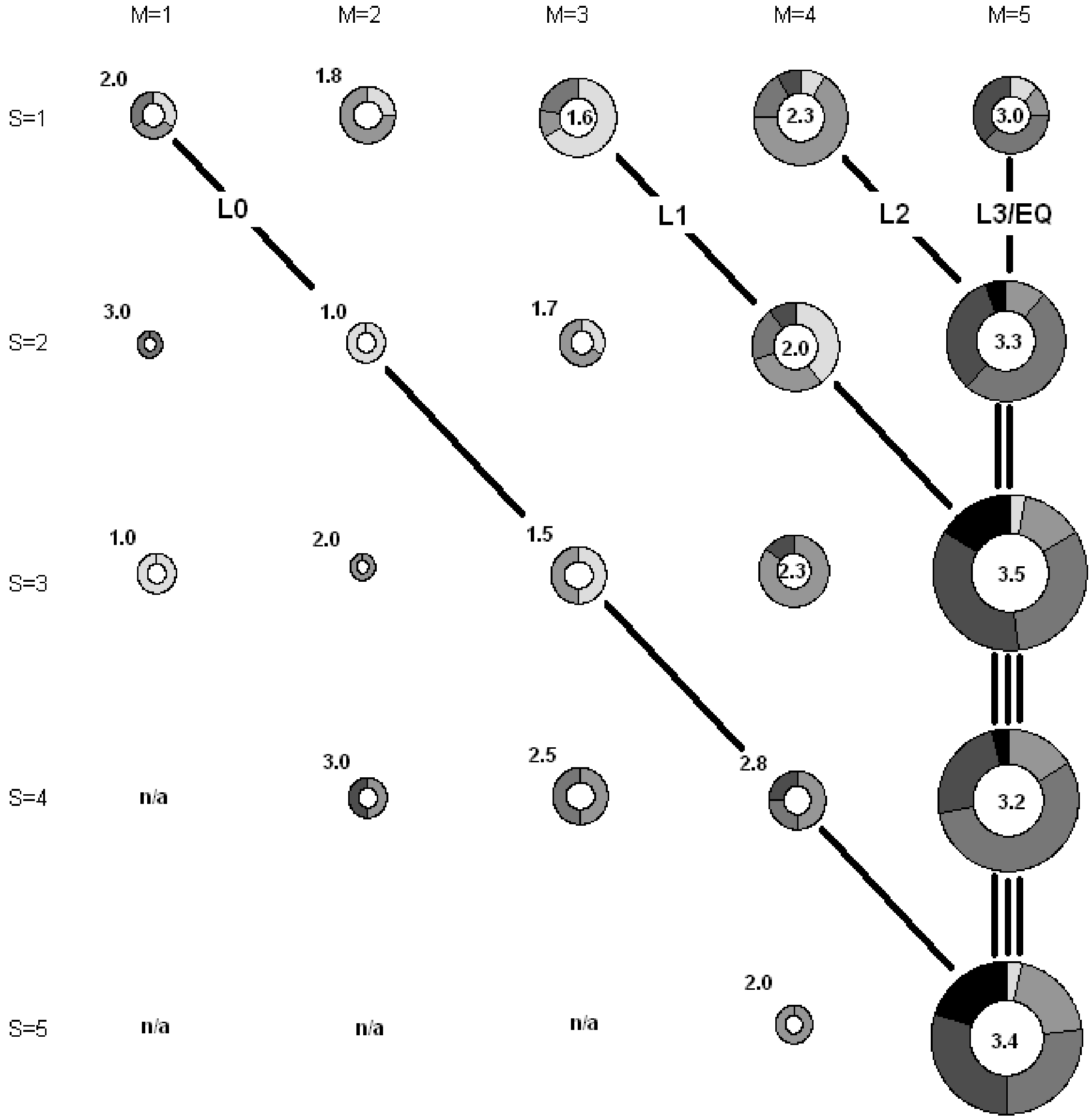
As b increases to 1 or 2, the amount of information transmitted decreases as predicted by Crawford and Sobel’s equilibrium comparative statics.

But there are also systematic deviations from the most informative (or from any) equilibrium, and lying and successful deception are common.

Figure 3 (next slide) shows the Sender’s message frequencies and the Receiver’s action frequencies as functions of the observed state S when b = 2.

In the essentially unique, most informative equilibrium when b = 2, M = {1, 2, 3, 4, 5}, so equilibrium message distributions would look the same for all five rows; and equilibrium actions would be concentrated on A = 3.

Figure 3: Raw Data Pie Chart (b=2) (Hidden Bias-Stranger)



However, although the observed actions are quite close to A = 3, the message distributions shift rightward as S increases (going down in the table). Thus:

- Most Senders exaggerate the truth (most messages are above the diagonal), apparently trying to move Receivers from Receivers' ideal action A = S toward Senders' ideal action A = S + 2 (or 5, if S + 2 > 5).

- Even so, there is some information in Senders' messages (the message distributions shift rightward going down in the table, so messages are positively correlated with the state).

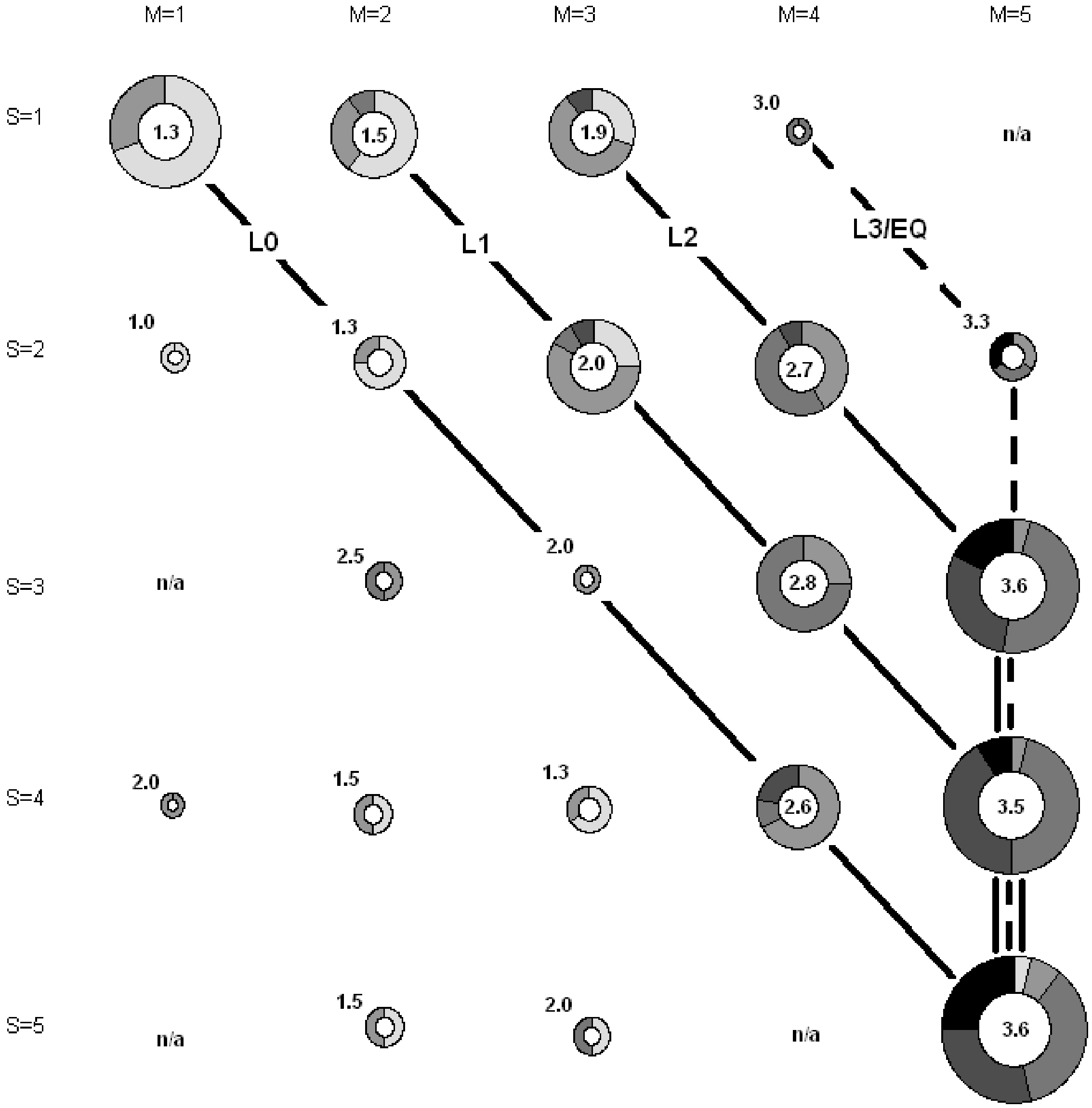
- Receivers are usually deceived to some extent (the average A is almost always > S).

Figure 2 (next slide) shows the Sender’s message frequencies and the Receiver’s action frequencies as functions of the observed state S when b = 1.

When b = 1, in the most informative robust equilibrium, the Sender’s message is M = 1 when S = 1 and M = {2, 3, 4, 5} when S = 2, 3, 4, or 5; and the Receiver chooses A = 1 when M = 1 and A = 3 or 4 when M = {2, 3, 4, 5}.

Thus, in equilibrium the distributions of messages and actions would be the same for S = 2, 3, 4, or 5.

Figure 2: Raw Data Pie Chart (b=1) (Hidden Bias-Stranger)



However, when b = 1:

- Senders again almost always exaggerate the truth (messages above the diagonal), apparently trying to move Receivers from Receivers' ideal action A = S toward Senders' ideal action A = S + 1 (or 5, if S + 1 > 5).

- Even so, there is again some information in Senders' messages (the message distributions shift rightward going down in the table, so the messages are positively correlated with the state).

- Receivers are again deceived to some extent (average A usually > S).

What kind of model can explain results like this? Wang et al., following Cai and Wang (2006 *GEB*), propose a level-*k* explanation in the style of Crawford's (2003 *AER*) analysis of preplay communication of intentions:

Anchor beliefs in a truthful Sender *L0*, which sets M = S; and a credulous Receiver *L0* (which also best responds to an *L0* Sender), setting A = M.

L1 Senders best respond to *L0* Receivers by inflating their messages by b: M = S + b (up to M = 5), so that *L0* Receivers will choose S + b, yielding the Sender's ideal action given S.

L1 Receivers (as defined by Wang et al.; the numbering is a convention) best respond to *L1* Senders by discounting the message, normally setting A = M – b, yielding Receivers' ideal action given M = S + b of S.

(The qualification "normally" reflects Wang et al.'s assumption that *L1* Receivers take into account that when b = 2, *L1* senders with S = 3, 4, or 5 all send M = 5, with the result that *L1* Receivers, knowing that S is equally likely to be 3, 4, or 5, choose A = 4 instead of A = M – 2b = 3.)

L2 Senders best respond to *L1* Receivers by inflating their messages by 2b: M = S + 2b (up to M = 5), so that *L1* Receivers will set A = M – b = S + b, yielding Senders' ideal action given S.

L2 Receivers best respond to *L2* Senders by discounting the message, normally setting A = M – 2b, yielding Receivers' ideal action given M = S + 2b of S.

(The qualification "normally" reflects Wang et al.'s assumption that *L2* Receivers take into account that when b = 1, *L2* senders with S = 3, 4, or 5 all send M = 5, with the result that *L2* Receivers, knowing that S is equally likely to be 3, 4, or 5, choose A = 4 instead of A = M – 2b = 3.)

L2 Receivers also take into account that when b = 2, *L2* senders with S = 2, 3, 4, or 5 send M = 5, so that *L2* Receivers, knowing that S is equally likely to be 2, 3, 4, or 5, choose A = 4 instead of A = M – 2b = 3.

Econometric estimation classifies 18% of 16 Sender subjects as *L0*, 25% *L1*, 25% *L2,* 14% *Sophisticated*, and 18% *Equilibrium* (note different type definitions).

Figures 2 and 3 show why.

(Note that when b = 1, *L1*, *L2*, and *Eq* all predict M = 5 when S = 4 or 5; and when b = 2, *L1*, *L2*, and *Eq* all predict M = 5 when S = 3, 4, or 5.)

Figure 2: Raw Data Pie Chart (b=1) (Hidden Bias-Stranger)

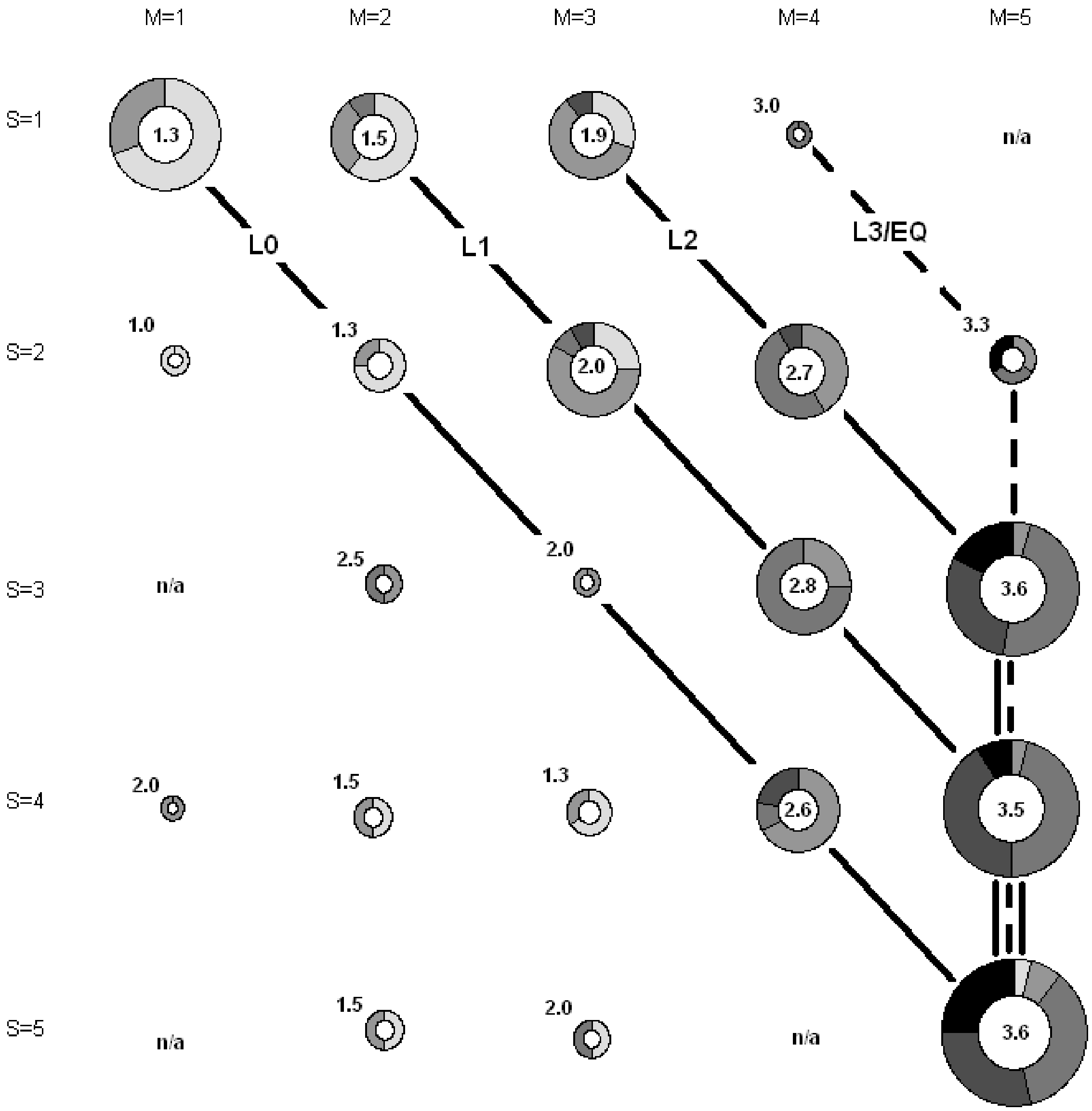
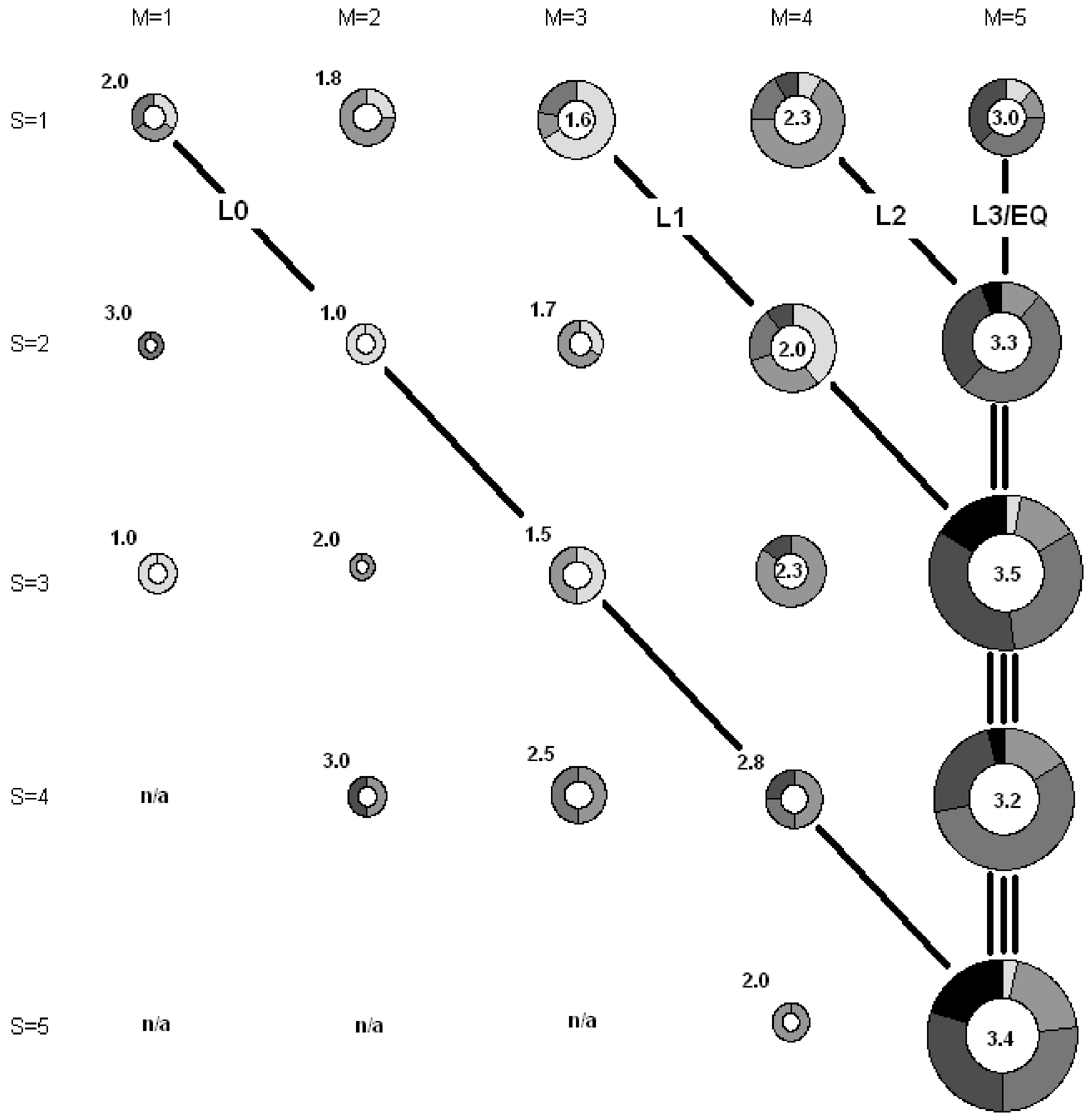


Figure 3: Raw Data Pie Chart (b=2) (Hidden Bias-Stranger)


M=1
M=2
M=3
M=4
M=5
S=1
S=2
S=3
S=4
S=5
L0
L1
L2
L3/EQ
n/a
n/a
n/a
n/a